UNIVERSITY OF HAWAII
LIBRARY
AUG 24 61

ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT of the PHILOSOPHICAL MAGAZINE

EDITOR
PROFESSOR B. H. FLOWERS, M.A., D.Sc.

CONSULTANT EDITOR
PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

EDITORIAL BOARD
SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.
SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.
PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

VOLUME 10 APRIL 1961 NUMBER 38

PRICE per part 25*s*. plus postage
PRICE per annum £4 15*s*. 0*d*. post free

PRINTED AND PUBLISHED BY TAYLOR & FRANCIS, LTD
RED LION COURT, FLEET ST., LONDON E.C.4

REVISED CHEAPER EDITION FOR LIBRARIES AND SCHOOLS

A History of Mathematics

From antiquity to the early nineteenth century

By J. F. SCOTT, B.A., D.Sc., Ph.D.

Vice-Principal of St. Mary's College, Strawberry Hill, Twickenham, Middlesex

Author of *The Scientific Work of René Descartes* (1596–1650), *Mathematical Work of John Wallis, D.D., F.R.S.* (1616–1703), and other works

CONTENTS: Mathematics in Antiquity—Greek Mathematics—The Invention of Trigonometry—Decline of Alexandrian Science and the Revival in Europe—Mathematics in the Orient—Progress of Mathematics during the Renaissance—New Methods in Geometry—The Rise of Mechanics—The Invention of Decimal Fractions and of Logarithms—Newton and the Calculus—Taylor and Maclaurin, the Bernoullis and Euler, Related Advances—The Calculus of Variations, Probability, Projective Geometry, Non-Euclidean Geometry—Theory of Numbers—Lagrange, Legendre, Laplace, Gauss. This volume is intended primarily to help students who desire to have a knowledge of the development of the subject but who have too little leisure to consult works and documents. The author has availed himself of the facilities afforded by the Royal Society and other learned Societies to reproduce extracts from manuscripts and many scarce works.

Size $9\frac{3}{4}'' \times 6\frac{3}{4}''$. 266 pp. **Price 27s. 6d.** plus postage and packing 2*s.* 0*d.*

Some Reviews of the First Edition

" The invention of trigonometry, decimal fractions, logarithms and the calculus are each discussed clearly and concisely. The book is easy to read for anybody who knows the elements of mathematics and, although not free from minor errors, can be strongly recommended."—*British Book News*, April 1958.

" Physicists will find that the development in applied mathematics are clearly set out, from ancient times, through that of Archimedes, to the mechanics of the sixteenth century when interest was revived. Significant advances made by Stevin, Galileo, Descartes, Huygens and others are stressed, and help the reader to appreciate what Newton achieved. There are useful appendices giving brief biographical notes on mathematical topics and terminology, followed by a bibliography."—*Proceedings of The Physical Society*, September 1958.

" . . . it has been written with clarity and balance, and the excellent printing helps to make it a pleasure to read."—*The Times Educational Supplement*, 21 March 1958.

" . . . his (Dr. Scott's) wide knowledge of the material, his careful description of methods combine to provide an account which at times gives a sense of the excitement of discovery."—*Nature*, 26 July 1958.

" The printers and publishers are to be congratulated upon having produced such an attractive volume, ... We feel sure the book will be received with delight by all those interested in mathematical histories."—*BEAMA Journal*, August 1958.

" The work comes to life mainly because of his admirable use of the writings of mathematicians themselves, which vividly illustrates the great difficulties under which many of them laboured. This is not a book for the layman but both the student and anyone to whom figures are a fascination, will find the subject clearly and pleasantly presented."—*Technical Bookguide*, March 1958.

Printed and Published by

TAYLOR & FRANCIS LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

An invitation to subscribe to
this Important New Pergamon International Research Journal

INFRARED PHYSICS

This journal is being established as an international research journal for the publication of scientific papers concerning infrared physics and its applications. It is concerned with infrared theory, experiment, and instrumentation as applied to infrared detection and transmission and to problems of atmospheric, meteorological, geophysical, astrophysical and space research. Except as they pertain directly to infrared studies of planetary and stellar atmospheres, papers on molecular spectroscopy or spectrochemical analysis are considered outside the scope of INFRARED PHYSICS. The journal will contain Research Papers, specially invited Critical Surveys, quickly published Research Notes, and Book Reviews. The language preferred is English, but papers will be published occasionally in French and German. Manuscripts for editorial consideration should be sent to the Member of the Board of Editors most conveniently located.

CONTENTS OF VOLUME 1 NUMBER 1 – JUST PUBLISHED

E. SCOTT BARR: The infrared pioneers-I. Sir William Herschel; N. C. BEESE: Light sources for optical communication; D. F. EDWARDS and M. MERCADO: Ultimate sensitivity and practical performance of the tellurium photoconductive detector; S. NEILSEN, W. D. LAWSON and A. F. FRAY: Some infrared transmitting glasses containing germanium dioxide; P. BRATT, W. ENGELER, H. LEVINSTEIN, A. MACRAE and J. PEHEK: A status report on infrared detectors; H. HAPP and L. GENZEL: Interferenz-Modulation mit monochromatischen Millimeter-Wellen; P. A. LAPP and H. S. KERR: Sunseeker for high-altitude infrared solar spectra; H. KALLMANN, J. RENNERT and M. SIDRAN: Infrared photography using persistent internal polarization in phosphor plates; C. HILSUM and P. R. HARDING: The theory of thermal imaging, and its application to the absorption-edge image-tube; R. BEER and J. RING: A. high-pressure scanning Fabry-Perot interferometer for the infrared; T. S. MOSS and A. G. PEACOCK: Infrared optical properties of lead halides (Research Note).

Published Quarterly

BOARD OF EDITORS

N. MIGEOTTE	Université de Liège, Institut d'Astrophysique, Cointe-Solessin, Belgium.
T. S. MOSS	Royal Aircraft Establishment, Radio Department, Ambarrow Court, Lower Sandhurst Road, near Camberley, Surrey, England.
S. PASSMAN	The RAND Corporation, 1700 Main Street, Santa Monica, California, U.S.A.
W. K. WEIHE	U. S. Army, Engineer Research and Development Laboratories, Fort Belvoir, Virginia, U.S.A.

Assisted by an International Honorary Editorial Advisory Board

A copy of the journal and details of subscription rates gladly sent on application

PERGAMON PRESS

OXFORD **LONDON** **NEW YORK**

Headington Hill Hall, Oxford 4 & 5 *Fitzroy Square, London W.*1

122 *East* 55*th Street, New York* 22, *N.Y.*

Radio Waves in the Ionosphere

K. G. BUDDEN

A full account of the mathematical basis of the theory of the propagation of radio waves in a horizontally stratified ionosphere. It is both a textbook for students, and those comparatively new to the subject, and a reference work for practising engineers and research workers in the field of radio communication. 95*s. net*

SOME STANDARD WORKS NOW AVAILABLE IN
PAPER BACK EDITIONS

The Mathematical Theory of Non-Uniform Gases

S. CHAPMAN & T. G. COWLING

17*s.* 6*d. net*

The Mathematical Theory of Electricity & Magnetism

SIR JAMES JEANS

25*s. net*

The Mathematical Theory of Relativity

A. S. EDDINGTON

16*s. net*

Spherical Astronomy

W. M. SMART

22*s.* 6*d. net*

The Evolution of Modern Physics

A. EINSTEIN & F. INFELD

15*s. net*

CAMBRIDGE UNIVERSITY PRESS

CONTENTS

The Theory of Impurity Conduction†

By N. F. Mott and W. D. Twose‡
Department of Physics, University of Cambridge

CONTENTS

§ 1. Introduction

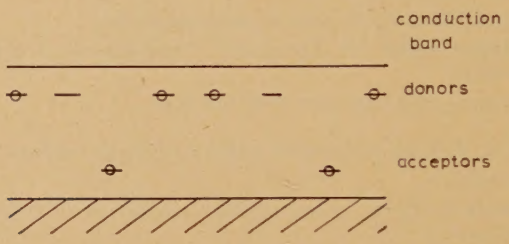
The purpose of this article is to study one of the ways in which electricity can flow in a semiconductor containing impurities. The current in an impure semiconductor is due to two competing conduction processes which act in parallel. The first process is responsible for the current usually observed which (in for example n-type material) is carried by electrons in the conduction band in thermal equilibrium with electrons on donor impurities. The second process arises as follows: An electron occupying an isolated donor has a wave function localized about the impurity and an energy slightly below the conduction band minimum. Because there is a small but finite overlap of the wave function of an electron on one donor with neighbouring donors, a conduction process is possible in certain circumstances in which the electron moves between centres by tunnel effect without activation into the conduction band. This we call *impurity*

† Supported in part by the U.S. Air Force.
‡ Now at the Institute for the Study of Metals, University of Chicago.

conduction. The electrons of high mobility in the conduction band completely dominate the conductivity at higher temperatures. However, although the mobility of an electron moving in the impurity levels is very small since it depends on interaction between widely spaced impurities, at low temperatures impurity conduction will dominate due to the absence of electrons in the conduction band.

The circumstance in which impurity conduction is possible is the presence of 'compensation', by which we mean the presence of some minority centres, acceptors in an n-type conductor. These accept electrons from a certain proportion of the donors, thus allowing the movement of electrons from an occupied donor into an occupied one (fig. 1). Without compensation impurity conduction is not possible, unless the overlap between the centres is very large; when this is large enough, corresponding to a critical concentration N_c, another form of conductivity sets in, in which the electrons behave like a degenerate electron gas. This we shall treat in part II of this report and shall describe as metallic impurity conduction.

Fig. 1



Energy diagram of an n-type semiconductor containing donors and acceptors. The horizontal lines represent centres, the circles electrons in them.

Impurity conduction† was first observed by Busch and Labhart (1946) in silicon carbide, and has since been observed in a large number of both n- and p-type semiconductors; references are given in § 4. The concept of an electron bound to a donor centre is complementary to that of a hole bound to an acceptor. Hence our discussion, which for convenience is based on n-type, can readily be carried over to p-type material. Most experimental work has been done on the valence semiconductors, germanium and silicon, and this article is mainly about these.

A feature of impurity conduction, which distinguishes it from the usual semiconduction, is its extreme sensitivity to impurity concentration. For example, a change by a factor of 30 in the density of impurities in germanium can alter the conductivity by the impurity process by a factor of 10^7 (fig. 4), while the corresponding change in the conductivity in the conduction band in the exhaustion range of temperatures (i.e. at those when the electrons are nearly all in the conduction band) is only of order 28. Another

† The possibility of impurity conduction was suggested on theoretical grounds by Schottky in 1935 (see Gudden and Schottky 1935).

feature is that, when the impurity concentration is small, the curves plotting $\ln \rho$ against $1/T$ exhibit a finite slope in the temperature range where impurity conduction predominates, suggesting that the charge transfer between impurity centres must itself be thermally activated. Above the critical concentration N_c mentioned above, the resistivity becomes independent of temperature; the conductivity is then apparently metallic and carriers move freely without thermal activation. There is a small transition region (a factor of order 4 in the impurity concentration in germanium) just below N_c, in which the conductivity is non-metallic but the slopes of the curves ($\ln \rho$ vs. $1/T$) decrease and finally vanish at the critical concentration. A complicated temperature dependence of the Hall effect is also observed in this region (fig. 5).

The theoretical interest of these phenomena is two-fold. First they give the opportunity of studying the transition from metallic to non-metallic conduction which occurs as the concentration of carriers is decreased. It has been postulated by one of us in a number of papers (Mott 1949, 1952, 1956, 1957, 1961) that as the lattice spacing of an *ordered* array of atoms is increased there should be a sharp transition from a metallic to a non-metallic state of the valence electrons. The theoretical treatment of this transition is a many-body problem, involving the interactions between the electrons. No way is known, except possibly the use of high pressures, of changing the interatomic distance of a crystalline array of atoms over a large enough range, so that to test this hypothesis we are driven back to a study of a disordered array of centres such as occurs in doped germanium or silicon.

The second point of interest is the mobility of an electron in a disordered lattice, considered as a one-body problem. For high concentrations giving metallic conductivity, we have a problem like that of a liquid metal, but with a greater degree of disorder. For low concentrations, it appears that the electron moves by a hopping process from one centre to another, interaction with phonons being essential and the concept of a mean free path not appropriate.

The article therefore divides naturally into two parts. In the first we shall be concerned with the experimental observation and calculation of impurity conductivity in the region of low concentration. We discuss in § 5 the reasons for using localized states and a phonon-activated hopping process in this low concentration region. The effect of compensating impurities and disorder is considered, and the theory of the interaction of localized carriers with lattice vibrations is traced through from the limits of strong coupling (polar semiconductors) to weak coupling (valence semiconductors). Although we describe this process as a one-body problem, we must in our applications of Fermi statistics introduce the interaction between electrons in the sense that an electron cannot move into an impurity centre that is already occupied. The second part of this article will have as its theme the interaction between carriers in the impurity centres when this becomes large enough to lead to a transition to a metallic form of

conductivity. This second section will start from an assumption of what is to be expected from a crystalline arrangement of centres and then discuss the effect of disorder and compensation.

Part I

§ 2. The Model and General Discussion of Impurity Conduction

If we neglect the dependence on direction of the effective mass, an electron occupying an isolated donor can be taken to move in a hydrogen-like orbit in the Coulomb field of the donor ion, with a Bohr radius

$$a_0 = \kappa (m/m^*) a_H$$

where m^* is the effective mass and a_H is the radius of a hydrogen atom (0·54 Å). Because of the large dielectric constant κ and small effective mass ratio m^*/m, the orbit may extend over several hundred of the host lattice sites. The energy of this state lies slightly below the lowest state of the conduction band. A similar description applies to a hole bound to a negative acceptor impurity; in this case the energy of the vacant state lies slightly above the top of the valence band. A more exact description of the impurity states is given in § 3, taking into account the dependence of m^* on direction.

In an n-type semiconductor (one in which the donor concentration N_D exceeds the accept or concentration N_A), at the absolute zero of temperature all the acceptors will be occupied and consequently negatively charged. The number of donors occupied and therefore neutral is $N_D - N_A$ (fig. 1). Overlap between wave functions corresponding to neighbouring sites allows movement from an occupied to an empty donor without activation into the conduction band. Our study of impurity conduction will therefore be a study of transport of electrons in a *random* lattice from one positively charged donor to another and in the field of fixed negatively charged acceptors. The host crystal is regarded as a dielectric medium in which this random impurity lattice is imbedded; thermal energy is supplied by vibrations of the host crystal.

As already emphasized in the introduction, at high concentrations of impurity the resistivity and Hall coefficient become independent of temperature at low temperatures, the electrons behaving like a degenerate electron gas. This behaviour, illustrated in figs. 4 and 5, is discussed in part II. In this section we consider low concentrations. There are two particularly simple cases which we may dicuss. If N_D is the concentration of majority centres (say donors) and N_A the concentration of minority centres (acceptors), these are:

Case (*a*). $N_D \gg N_D - N_A$

There will then be a *small* number $(N_D - N_A)$ of electrons in the donor states, and, owing to the random arrangement of immobile positively and negatively charged centres, a random fluctuation in the potential

energy from one centre to another. The question then arises, as in all these considerations (§ 5), whether in the absence of lattice vibrations the characteristic wave functions (solutions of the Schrödinger equation for a single electron) are localized or whether they spread through the lattice. By a localized wave function we mean one that decays exponentially to zero at large enough distances from a given point in space. Considerations set out in § 7 show that in a one-dimensional lattice they are always localized. In a three-dimensional lattice they are localized if the degree of disorder, or the ratio of the energy in the random field to the band width, are great enough (§ 5). We believe this to be the case in the range of concentration for which experimental measurements are made; the electron can then jump from one centre to another only with the help of *phonons*. The process by which it does so is a main theme of this report. But it is possible that, as the concentration of donors increases, there may be formed unbound states so that the activation energy for motion would be zero. It is possible that a 'crystallization' of electrons, as envisaged by Wigner (1938) may occur, and a transition to a metallic state (condensed electron gas) only for higher concentrations of electrons. This is discussed in part II and by Mott (1961).

Case (*b*). $N_D \gg N_A$

In this case most of the donors are occupied and a small number vacant. The donor states which are unoccupied are to be thought of as carriers; as electrons jump from occupied to unoccupied donors, the positive vacancy moves through the lattice. As in case (*a*) when the centres are still localized, we think of this as a hopping process from centre to centre.

In case (*b*) it is possible to discuss the activation energy for charge transfer in terms of a simple model first suggested by one of us (Mott 1956). The carrier (the positive charge vacancy) will in its state of lowest energy lie as closely as possible to a negatively charged acceptor. Before conduction can occur the carrier must be thermally activated from this bound state. An energy of order

$$E = (e^2/\kappa)\langle 1/r_{AD}\rangle = 1{\cdot}46\,(e^2/\kappa)\,N_D^{1/3} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

will be required to remove the carrier from the neighbourhood of the acceptor. Here r_{AD} is the nearest neighbour separation between a donor and an acceptor, and $\langle 1/r_{AD}\rangle$ denotes an average, assuming a Poisson distribution of centres. Price (1957) suggested as a better approximation that E should be of order

$$E \sim (e^2/\kappa)(N_D^{1/3} - 2N_A^{1/3}); \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.1)$$

he supposed that one should not consider removing the carrier to infinity but to a distance at which it is effectively outside the field of the particular acceptor concerned. The further movement of the 'free' carriers through the lattice may well require further activation, but the energies will be considerably smaller. Price has examined the statistics of this model.

We give here a simple derivation of the number n of free carriers. Assuming there is only one trap site associated with each acceptor, and a constant trapping energy E, the free energy of the carriers is:

$$nE - kT\,[\ln\{N_D!/n!(N_D-n)!\} - \ln\{N_A!/n!(N_A-n)!\}].$$

Minimizing this with respect to n, we obtain

$$\left.\begin{aligned} n^2/(N_D-n)(N_A-n) &= \exp(-E/kT) \\ \text{or} \qquad n &\simeq (N_D N_A)^{1/2}\exp(-E/2kT) \end{aligned}\right\} \quad . \; . \; . \quad (2)$$

at low temperatures. Thus we expect the slope ϵ_3 of the curve plotting $\ln\rho$ against $1/kT$ to be approximately $\frac{1}{2}E$ at low temperatures. Saturation in the carrier concentration should set in at temperatures given by

$$\exp(-E/2kT) \sim K^{1/2}, \quad (K = N_A/N_D). \quad . \; . \; . \; . \quad (3)$$

In, for example, p-type germanium containing $6{\cdot}5 \times 10^{15}$ majority impurities per cm^3 and a compensation ratio $K = 0{\cdot}05$, the observed activation energy at low temperatures is $1{\cdot}6 \times 10^{-3}$ ev. Therefore the saturation temperature T_s should be approximately 14°. A slight flattening of the resistivity curve is observed at around 5°, the high temperature end of the impurity conduction range (fig. 9, specimen (*a*)). However, this saturation effect is prominent only in specimens in which both the compensation and impurity concentration are small. Also, T_s does not follow in detail the prediction of (3). For example, in specimen (*b*) of fig. 9 (for which $K \sim 0{\cdot}4$) T_s should be 11°, lower than in specimen (*a*), whereas the observed slope is practically constant. This is one example of the limitations of the trapping model. We emphasize again that the model is valid only for very small K, when the acceptors are separated in general by a large number of donors.

The theory of Miller and Abrahams (1960), outlined in § 8, is not limited to small K and thus cannot use the assumption of trap sites. Moreover it is essential in their work to remember that some sites are occupied and some not, so that the electron or 'hole' cannot move except where there is a site ready for it. These authors find that the resistivity ρ is given by

$$\rho(T) \propto \exp(\epsilon_3/kT)$$

where

$$\epsilon_3 = \zeta - 1{\cdot}35\,\epsilon_A$$

and

$$\epsilon_A = (e^2/\kappa)\,(4\pi N_A/3)^{1/3}$$

and ζ is the Fermi energy. This is defined, if

$$f_i = 1/[1 + \exp\{(\epsilon_i - \zeta)/kT\}]$$

and ϵ_i is the energy, due to the random field, of the donor site i, by

$$\sum f_i = N_D - N_A.$$

The summation is approximated by an integration using a density of states function $p(\epsilon)\,d\epsilon$. This is obtained by assuming the energy spread to arise from nearest neighbour negatively charged acceptors. The

probability that an acceptor is at a distance r from a donor, and is the nearest one, is

$$p(r)\,dr = \frac{3r^2}{{r_A}^3}\exp\left\{-\left(\frac{r}{r_A}\right)^3\right\}dr,$$

where the mean acceptor separation is given by

$$r_A = (3/4\pi N_A)^{1/3}.$$

Then $\epsilon = e^2/\kappa r$, so we have

$$p(\epsilon)\,d\epsilon = (3{\epsilon_A}^3/\epsilon^4)\exp\{-(\epsilon_A/\epsilon)^3\}\,d\epsilon,$$

where

$$\epsilon_A = \epsilon^2/\kappa r_A.$$

On performing the integration of f_i over ϵ, we find

$$1-K = \exp\{-(\epsilon_A/\zeta)^3\}\{1+\exp(-\zeta/kT)\}]^{-1}$$

which determines ζ. Unless K is extremely small or very close to unity,

$$\zeta = -\epsilon_A\{\ln(1-K)\}^{-1/3}.$$

As we shall see in § 4, the magnitude of the resulting activation energy ϵ_3 agrees well with experimental values (cf. fig. 6). For $K \lesssim 0{\cdot}2$,

$$\epsilon_3 = \epsilon_D - 1{\cdot}35\,\epsilon_A = 1{\cdot}61\,(e^2/\kappa)(N_D{}^{1/3} - 1{\cdot}35N_A{}^{1/3}), \quad . \quad . \quad . \quad (4)$$

which is similar in form to that predicted by Price (1957), formula (1.1).

Finally we may ask whether, in case (*b*), we may expect for holes a phenomenon similar to that suggested in case (*a*) where the overlap becomes large, namely a state of affairs when the states for the hole are not localized. It is clear that, if $K \ll 1$, the lowest state must be localized with an energy given by (1), because a Coulomb field *always* leads to bound states. However, when the carrier has escaped from the field of the nearest charged acceptor, then for a high concentration of centres a hopping process may no longer be an appropriate description of the motion.

In the remainder of this section we shall suppose that K is less than $\frac{1}{2}$. The temperatures at which impurity conduction can be observed is thus determined by the following factors. The mobility of a carrier moving in the impurity levels is much smaller than in the conduction band, since the former is determined by interactions between widely spaced impurities. On the other hand, the number of carriers in the conduction band is determined by an activation energy ϵ_1 ($\sim 10^{-2}$ ev for germanium), while the energy ϵ_3 regulating impurity conduction is at least an order of magnitude smaller, varying from zero to 10^{-3} ev, depending on the concentration of centres. Thus at higher temperatures the conductivity is determined by carriers in the conduction band, at low temperatures by those in the impurity levels, and the transition between the two regions is quite sharp (fig. 4). In the whole region we may write formally

$$\sigma = n_c e\mu_c + ne\mu$$

where n_c, μ_c are the number of carriers and the drift mobility in the conduction band and n, μ the same quantities for the impurity carriers.

The Hall coefficient R, in the same notation, should depend on these quantities according to the expression

$$R=(n_c\mu_c\mu_{Hc}+n\mu\mu_H)/ec(n_c\mu_c+n\mu)^2,$$

where μ_{Hc}, μ_H is the Hall mobility of the conduction band and impurity electrons respectively. Neglecting any temperature variation in the mobilities, this expression has a maximum when

$$n_c e\mu_c=ne\mu. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The Hall curves for high concentration (samples 14, 15, fig. 5) can be qualitatively explained by assuming that a Hall effect exists for carriers in impurity levels, with Hall mobility comparable to the drift mobility. At high temperatures, for which $n_c\mu_c \gg n\mu$, so that the normal conduction band Hall effect is observed,

$$R_c=\frac{1}{ec}\left(\frac{\mu_{Hc}}{\mu_c}\right)\frac{1}{n_c}.$$

At temperatures below the Hall maximum, where $n_c\mu_c \ll n\mu$

$$R_{imp}=\frac{1}{ec}\left(\frac{\mu_H}{\mu}\right)\frac{1}{n}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

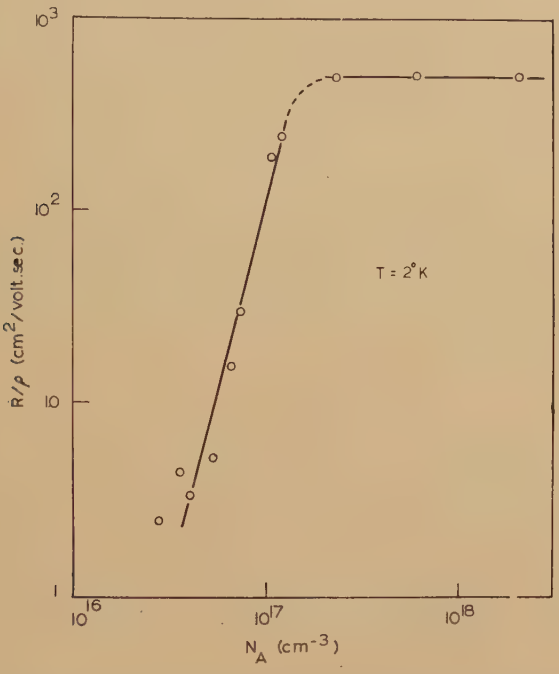
The number of carriers is practically independent of temperature in this range ($n \simeq N_D-N_A$), giving a temperature-independent Hall curve, as observed (fig. 5). The Hall coefficient R_{exh} in the exhaustion range (temperatures between 77° to 300°K) is due to the same number (N_D-N_A) of carriers, in the conduction band, but measured values of R_{exh} are about eight times larger than R_{imp}. This suggests that μ_H/μ is anomalously low. This is at present unexplained.

For the samples of low concentration it is not known whether a Hall effect exists at the impurity conduction temperatures. The drop in the Hall curves beyond the maximum can be attributed quantitatively to the rapidly decreasing contribution of the conduction band current to the total current. A flattening of the curves beyond the maximum (which would indicate a finite Hall mobility) is not observed down to the lowest temperatures at which it has been possible to make measurements (fig. 5). It is not clear on theoretical grounds whether a Hall effect is to be expected when the conduction process involves jumps of bound carriers to neighbouring sites. The experimental evidence suggests that R may exist, but be too small to be measured in the low concentration range. This is shown by the concentration dependence of R/ρ (fig. 2), measured at 2°K for the p-type samples of table 1. We see that R/ρ (which from (6) is proportional to μ_H) becomes small as the concentration falls through the transition region. This is discussed further in part II.

When charge carriers are positive vacancies (in n-type material with both impurity concentration and degree of compensation small), it might be thought that if a Hall effect exists it should be of opposite sign to that due to electrons in the conduction band. However, in a magnetic field H the vacancies do not behave as true particles with positive effective mass

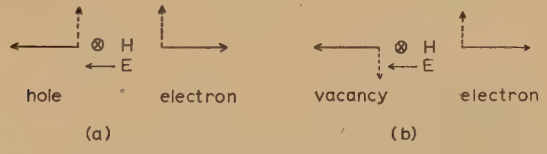
and charge. For example, if a positive and a negative particle are initially drifting in opposite directions, a magnetic field will deflect both particles in the same direction. However, a charge vacancy always moves in the opposite direction to an electron, even in a magnetic field (fig. 3). Hence we expect no change in sign in the Hall effect when the conduction is by charge transfer†.

Fig. 2



The Hall mobility R/ρ for p-type germanium with 40% compensation plotted against concentration of majority carriers (sample particulars in table 1).

Fig. 3



The motion in a crossed electric field E and magnetic field H of (*a*) an electron and a hole, (*b*) an electron and a charge vacancy (in a weakly compensated semiconductor).

† Yonemitsu *et al.* (1960) recently reported observing a change in sign of the Hall effect in a p-type germanium specimen containing 2×10^{16} gallium impurities per cm^3 and 40% compensation. However, although the magnitudes of their resistivity and Hall measurements agree well with those of Fritzsche (figs. 4, 5), the latter author observed no sign change.

In the band theory of periodic lattices the concept of a 'hole' which behaves as a positive charge with positive effective mass depends in part on the negative effective mass of electrons occupying states near the band maximum. This latter is a consequence of the increasing importance of Bragg reflection of electrons as their energies approach the band maximum. It is unlikely that the phases of electrons scattered by a completely disordered lattice will match sufficiently well for electrons of any energy to show a negative effective mass. Hence, in the metallic region of impurity conduction in n-type material, we would not expect to observe hole conduction or a positive Hall effect, for any concentration on impurity electrons. It is for this reason among others that we have avoided use of the misleading term 'impurity-band' when speaking of the metallic region.

§ 3. The Impurity Wave Functions

We have assumed in the last section a hydrogen atom model for an isolated impurity centre. The departures from this model are briefly outlined here. It has been shown by Kohn and Luttinger (see Kohn 1957) that the donor electron wave functions in germanium and silicon have the form

$$\psi^{(i)} = \sum_{j=1}^{N} \alpha_j^{(i)} F_j(\mathbf{r})\, \phi_j(\mathbf{r}), \qquad (7)$$

where ϕ_j is the Bloch wave function at one of the conduction band minimum denoted by j, and the sum is over the N equivalent minima ($N=4$ for germanium, $N=6$ for silicon). The α_j are coefficients which are determined by the symmetry of the state. $F_j(\mathbf{r})$ is an envelope function satisfying a Schrödinger equation for the potential due to the impurity, but with the free electron mass replaced by the effective mass appropriate to the jth minimum:

$$\left\{ -\frac{\hbar^2}{2m_l}\frac{\partial^2}{\partial z^2} - \frac{\hbar^2}{2m_t}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - \frac{e^2}{\kappa r} \right\} F_j(\mathbf{r}) = \epsilon_1 F_j(\mathbf{r}).$$

The z-axis lies along the direction of the k-vector of the jth minimum, and m_l, m_t are the longitudinal and transverse masses respectively. For the ground state, F_j has the hydrogen-like form

$$F_j(\mathbf{r}) = (\pi a^2 b)^{-1/2} \exp\left\{ -\left(\frac{x^2+y^2}{a^2} + \frac{z^2}{b^2} \right)^{1/2} \right\}. \qquad (8)$$

For germanium, $a = 64{\cdot}5$ Å, $b = 22{\cdot}7$ Å. Corrections to this effective-mass approximation partially lift the N-fold degeneracy of the ground state. These corrections arise chiefly from departures from a simple Coulomb potential in the immediate neighbourhood of the impurity ion, and because the concept of a uniform dielectric constant breaks down in that region. In germanium, the ground state splits into a non-degenerate lower level and a 3-fold degenerate upper level (Price 1956); the coefficients of the wave functions (7) corresponding to these levels are:

$$\left.\begin{aligned}\alpha^{(1)} &= \tfrac{1}{2}(1,\ \ 1,\ \ 1,\ \ 1) \\ \alpha^{(2)} &= \tfrac{1}{2}(1,-1,\ \ 1,-1) \\ \alpha^{(3)} &= \tfrac{1}{2}(1,\ \ 1,-1,-1) \\ \alpha^{(4)} &= \tfrac{1}{2}(1,-1,-1,\ \ 1)\end{aligned}\right\}\quad \begin{aligned}&\text{lower level;}\\ &\\ &\text{upper level.}\end{aligned}$$

The splitting is $0{\cdot}57 \times 10^{-3}$ ev for antimony impurities, but is of order ten times larger for arsenic and phosphorus donors (Fritzsche 1960 a); hence the upper states can be neglected except in antimony.

The acceptor states can similarly be constructed from Bloch orbitals at the valence band maximum. In this case the envelope functions F satisfy a set of six coupled effective-mass equations (Kohn and Luttinger 1955, Kohn and Schechter 1959); the solution of these is difficult and hence only approximate solutions of the acceptor wave functions are known.

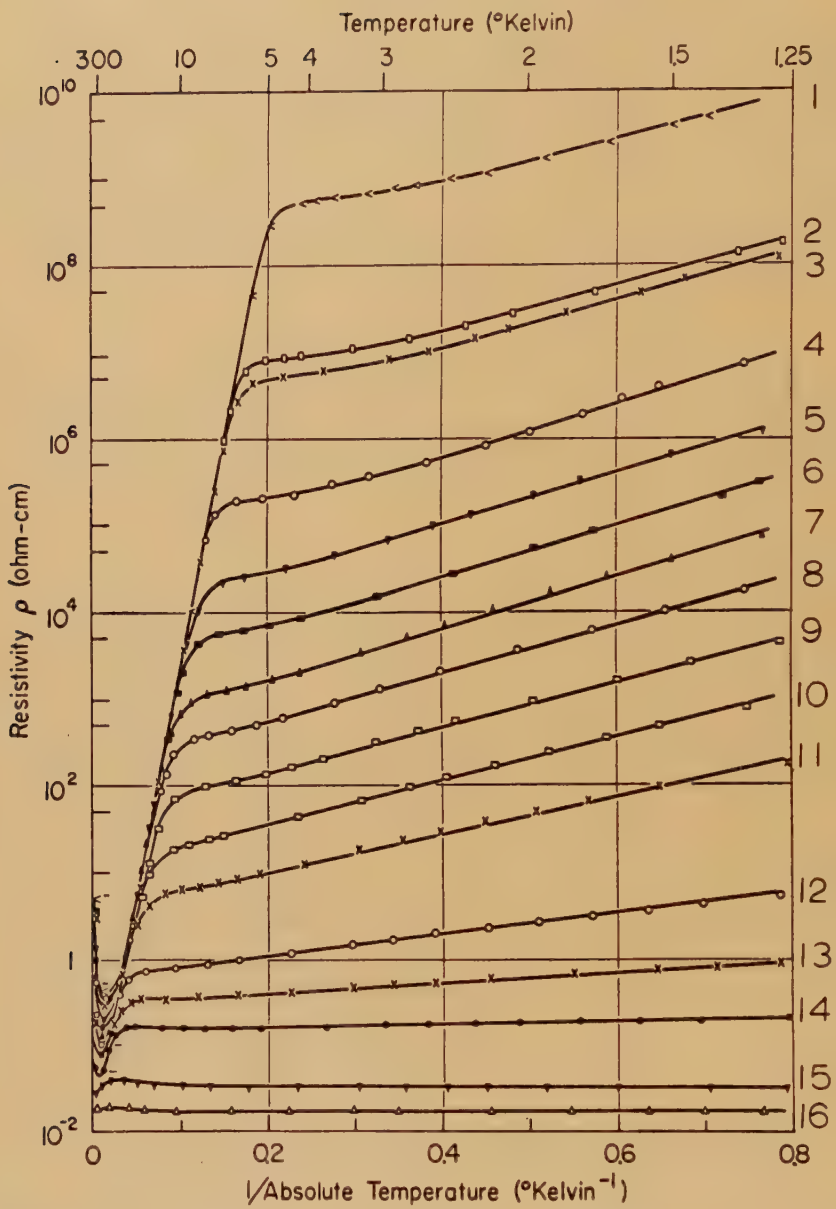
§ 4. Observations of Impurity Conduction

Impurity conduction has been observed in many semiconductors at low temperatures†. We do not attempt to give here an exhaustive review of experimental results but confine ourselves to the case of germanium, which has been studied most extensively. Results for other materials are qualitatively similar.

Prior to 1960, measurements in germanium and silicon were made in samples grown from melts doped with suitable impurities. For a quantitative comparison of theory and experiment it is desirable to have a range of measurements on samples where either (*a*) the majority impurity concentration is varied, but the degree of compensation K is kept constant, or (*b*) K is varied but the majority concentration is kept constant. It was practically impossible to achieve these conditions in the early measurements. Consequently, since the resistivity and Hall coefficient vary extremely rapidly with impurity concentration, comparison of theory and experiment was uncertain. Recently, however, Fritzsche and Cuevas (1960 a) has published measurements on p-type germanium samples in which the acceptor concentration (gallium) ranges between $8 \times 10^{14}\,\mathrm{cm}^{-3}$ to $1{\cdot}3 \times 10^{18}\ \mathrm{cm}^{-3}$ and the compensation ratio (arsenic and selenium) is kept fixed at $K = 0{\cdot}4$. The impurities were introduced into pure germanium by slow neutron bombardment (Cleland *et al.* 1950), causing transmutation of germanium atoms. The proportion of different impurities produced is determined by the cross sections for neutron capture and the decay schemes of the various germanium isotopes; therefore the compensation ratio is constant. The magnitude of the impurity concentration can be varied by the neutron flux and exposure times of different samples. Figures 4 and 5 show results of measurements of the resistivity ρ and Hall coefficient R; table 1 gives information about the specimens.

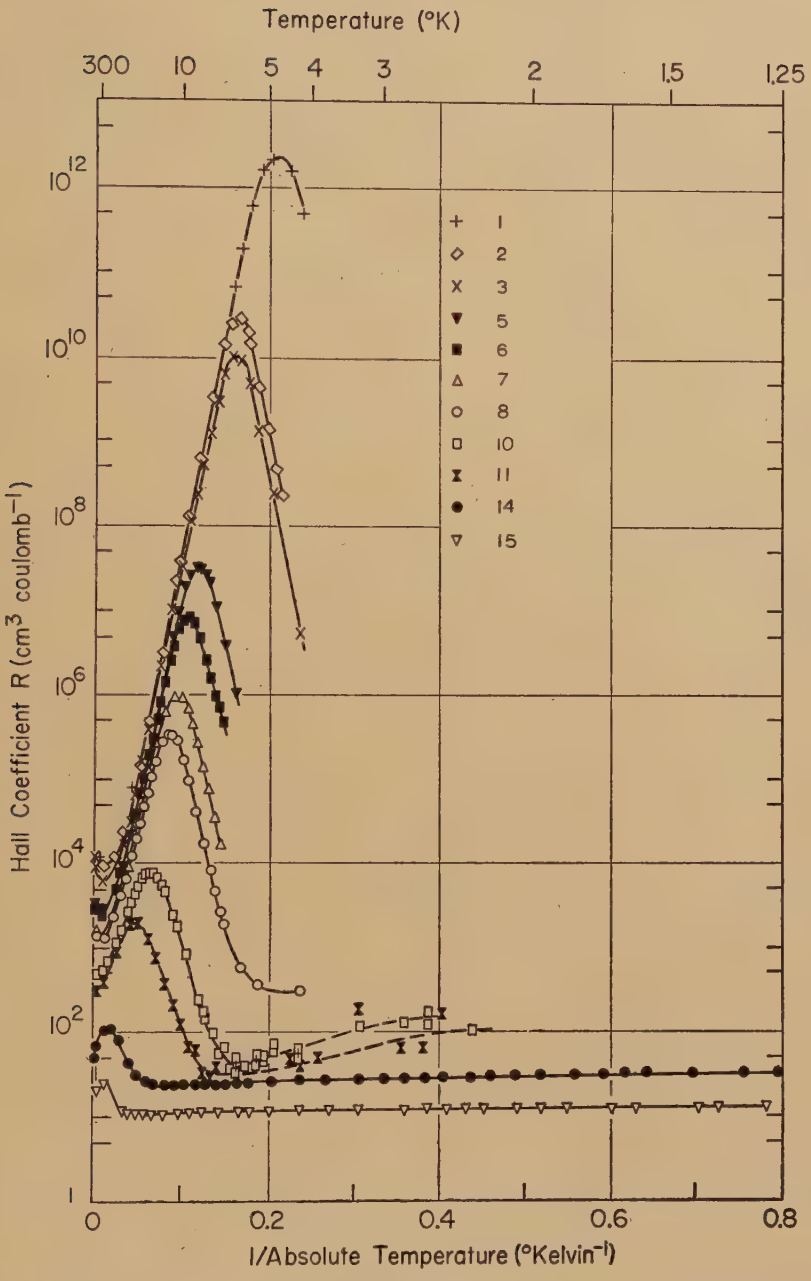
† SiC: Busch and Labhart (1946); Ge: Hung and Gliessman (1950, 1954); CdS: Kroger *et al.* (1954); Si: Morin and Maita (1954), Carlson (1955); p-InSb: Fritzsche and Lark-Horovitz (1955); n-InSb: Sladek (1958); Te: Fukuroi *et al.* (1954).

Fig. 4



Resistivity as a function of temperature of p-type germanium with compensation $K = 0{\cdot}4$; particulars in table 1 (Fritzsche and Cuevas 1960 a).

Fig. 5



Hall constants of p-type germanium as functions of temperature (particulars in table 1).

Table 1

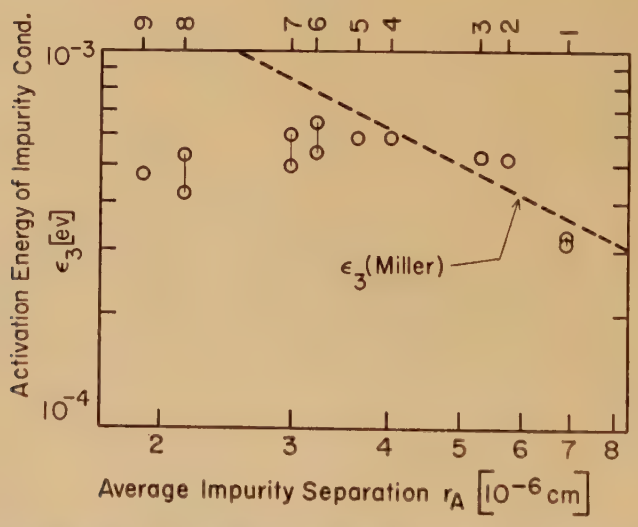
Sample	N_A (cm^{-3})	$\rho(T=2{\cdot}5°K)$ Ω cm	μ(cm^2/volt sec)
1	$7{\cdot}5\times10^{14}$	$8{\cdot}9\times10^{9}$	$4{\cdot}3\times10^{-5}$
2	$1{\cdot}4\times10^{15}$	$1{\cdot}8\times10^{7}$	$1{\cdot}2\times10^{-3}$
3	$1{\cdot}5\times10^{15}$	$1{\cdot}0\times10^{7}$	$2{\cdot}3\times10^{-3}$
4	$2{\cdot}66\times10^{15}$	$5{\cdot}6\times10^{5}$	$3{\cdot}0\times10^{-2}$
5	$3{\cdot}6\times10^{15}$	$1{\cdot}0\times10^{5}$	0·11
6	$4{\cdot}9\times10^{15}$	$2{\cdot}5\times10^{4}$	0·53
7	$7{\cdot}2\times10^{15}$	$6{\cdot}3\times10^{3}$	2·4
8	$9{\cdot}0\times10^{15}$	$2{\cdot}0\times10^{3}$	
9	$1{\cdot}4\times10^{16}$	$4{\cdot}5\times10^{2}$	
10	$2{\cdot}4\times10^{16}$	$1{\cdot}0\times10^{2}$	1·6
11	$3{\cdot}5\times10^{16}$	28	4·0
12	$7{\cdot}3\times10^{16}$	20	
13	$1{\cdot}0\times10^{17}$	0·50	
14	$1{\cdot}5\times10^{17}$	0·18	180
15	$5{\cdot}0\times10^{17}$	$3{\cdot}2\times10^{-2}$	250
16	$1{\cdot}35\times10^{18}$	$1{\cdot}8\times10^{-2}$	

Estimated mobilities μ are shown in column 4 of table 1.

For specimens (1–7) with low impurity content, μ is obtained from values of the resistivity at the temperature of the Hall maximum by making use of eqn. (5):

$$\mu=\mu_c n_c/n=(N_A K e\rho_c)^{-1}.$$

Fig. 6



The activation energy ϵ_3 of impurity conduction for the samples in table 1. The dashed curve represents the calculation of Miller and Abrahams (1960).

We have assumed the number of impurity carriers n to be $N_D (= KN_A)$, since K is less than $\frac{1}{2}$. At higher concentrations, we use simply

$$\mu = R/\rho.$$

This is really the Hall mobility, which as we have seen in § 2 may be appreciably smaller than the drift mobility.

Values of the activation energy ϵ_3 in the low temperature region are plotted in fig. 6, against the average acceptor separation. The magnitude of ϵ_3 agrees well with the values calculated from the theory of Miller and Abrahams (1960), shown by the dotted line, in the region of low concentration.

Measurements of the conductivities $1/\rho$ of n-type samples can be fitted by a sum of three exponentials (see, for example, Fritzsche 1958),

$$1/\rho = c_1 \exp(-\epsilon_1/kT) + c_2 \exp(-\epsilon_2/kT) + c_3 \exp(-\epsilon_3/kT).$$

Here ϵ_1 is the activation energy for exciting an electron into the conduction band, and ϵ_3 that for impurity conduction. The role of ϵ_2, which occurs only for samples in the transition region ($2 \times 10^{16} < N < 8 \times 10^{16}$), is not clear. ϵ_2 is observed in weakly compensated n-type and p-type samples, not however in p-type samples having $K = 0{\cdot}4$. A fourth activation energy was observed at temperatures below 1°K (Zaravstikaya 1956), in low concentration specimens. This however has since been shown to be due to stray light quanta exciting electrons into the conduction band. The resistivity is so large at these temperatures (of order $10^{10}\,\Omega\,.\,\mathrm{cm}$) that a very small fraction of excited electrons can lower the resistance appreciably.

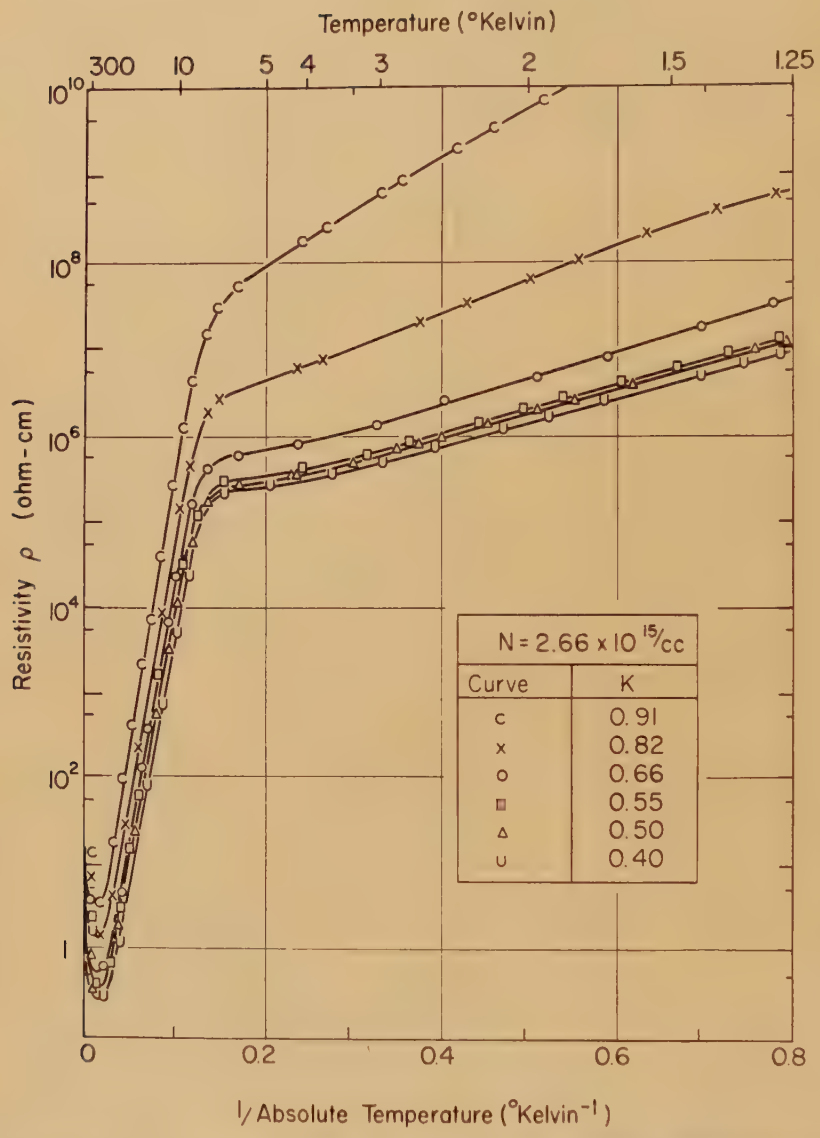
Some measurements of the resistivity have recently been reported on germanium samples of constant acceptor concentration ($2{\cdot}66 \times 10^{15}\,\mathrm{cm}^{-3}$) in which the degree of compensation was varied from 0·4 to 0·9 (Fritzsche and Cuevas 1960 b). These are shown in fig. 7.

The samples were produced by bombarding specimens of n-type germanium with slow neutrons. The same neutron flux and exposure time was used, so that the same acceptor (gallium) concentration was produced in each specimen. The degree of compensation however depends on the initial donor concentration. Figure 8 shows the observed activation energy as a function of K. The solid line is that predicted by Miller and Abrahams (1960) and is in good agreement, except at very high values of K.

The variation of the resistivity with the degree of compensaton should depend on whether the specimen shows metallic or non-metallic conduction. Measurements (in addition to those above) in which K is varied but the majority concentration is kept constant, have been made by Fritzsche (1955), Fritzsche and Lark-Horovitz (1959), in n- and p-type germanium, and by Ray and Longo (1959) in n- and p-type silicon. The results can be summarized as follows.

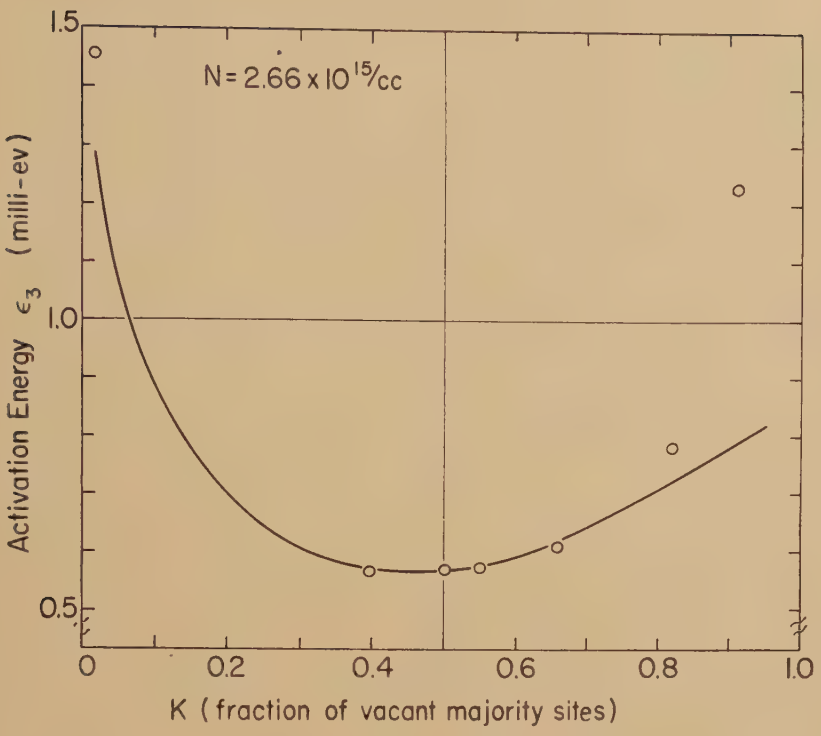
(i) In specimens showing metallic conductivity when $K \sim 0$, both the resistivity and Hall coefficient increase with K, the activation energy ϵ_3 remaining zero provided K is not too large. In p-type germanium samples

Fig. 7



The resistivity of p-type germanium with constant acceptor concentration N_A and compensation K varying from 0·4 to 0·9.

Fig. 8



with $N_A = 2{\cdot}5 \times 10^{17}\,\mathrm{cm}^{-3}$, Fritzsche and Lark-Horovitz (1959) observed that the conductivity becomes non-metallic as K increased between 0·4 and 0·7, the Hall and resistivity curves showing behaviour types of the transition region (figs. 9, 10).

(ii) In specimens showing non-metallic conductivity when weakly compensated, the resistivity at a constant temperature decreases to a minimum as K is increased to about 0·4, and thereafter increases.

Writing

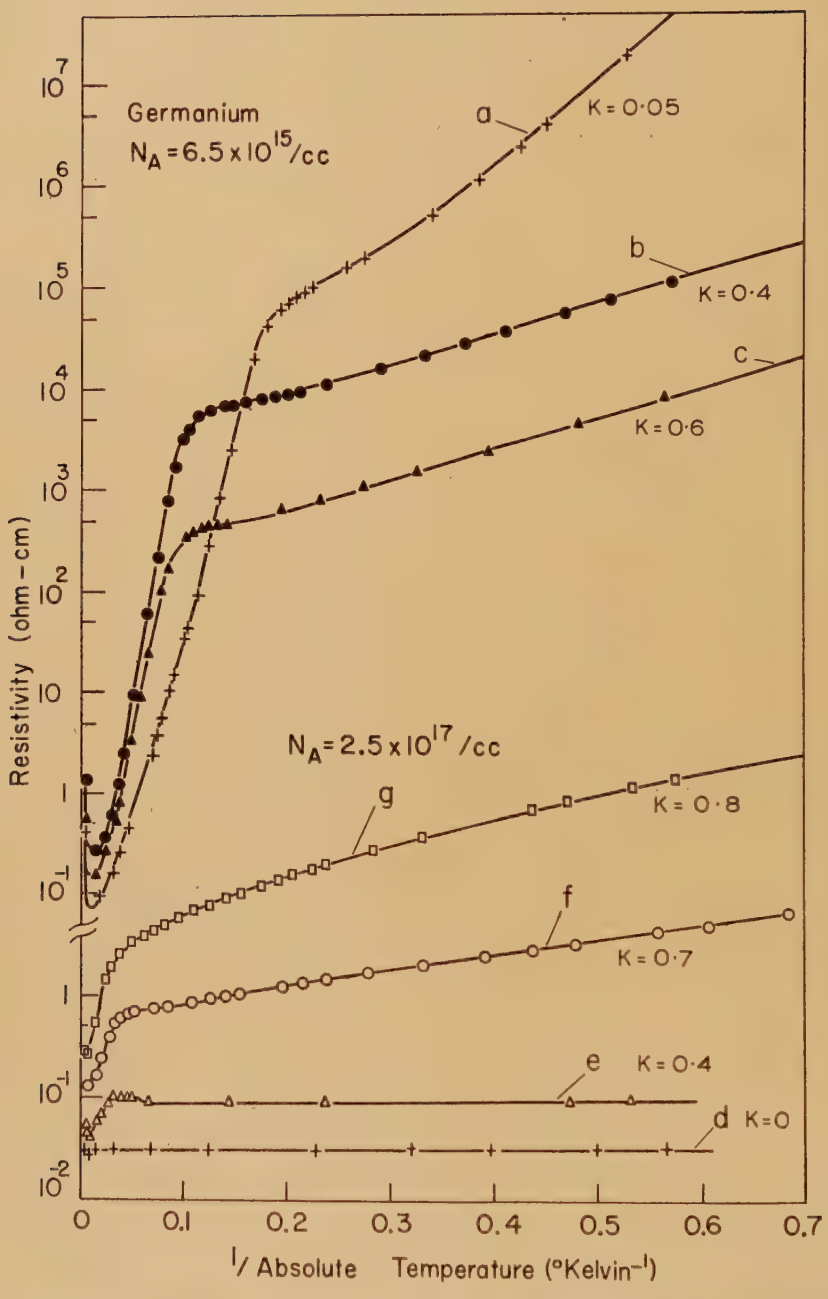
$$\rho = \rho_0 \exp(-\epsilon_3/kT),$$

ρ_0 is found to increase with K (fig. 11). However, in samples with $N_D = 10^{16}\,\mathrm{cm}^{-3}$, ρ_0 remains approximately constant in the range $10^{-3} < K < 10^{-2}$ (Fritzsche 1960, private communication); these samples have a non-vanishing Hall coefficient and are in the transition range.

These results can be understood in part by considering the effect of compensation on the carrier concentration. In the metallic range the number n of carriers is $N_{\mathrm{maj}}(1-K)$. In the non-metallic range $n = N_{\mathrm{maj}} K$ when $K < \frac{1}{2}$ (the carriers are vacancies on impurity sites) and $n = N_{\mathrm{maj}}(1-K)$ when $K > \frac{1}{2}$. Hence, since

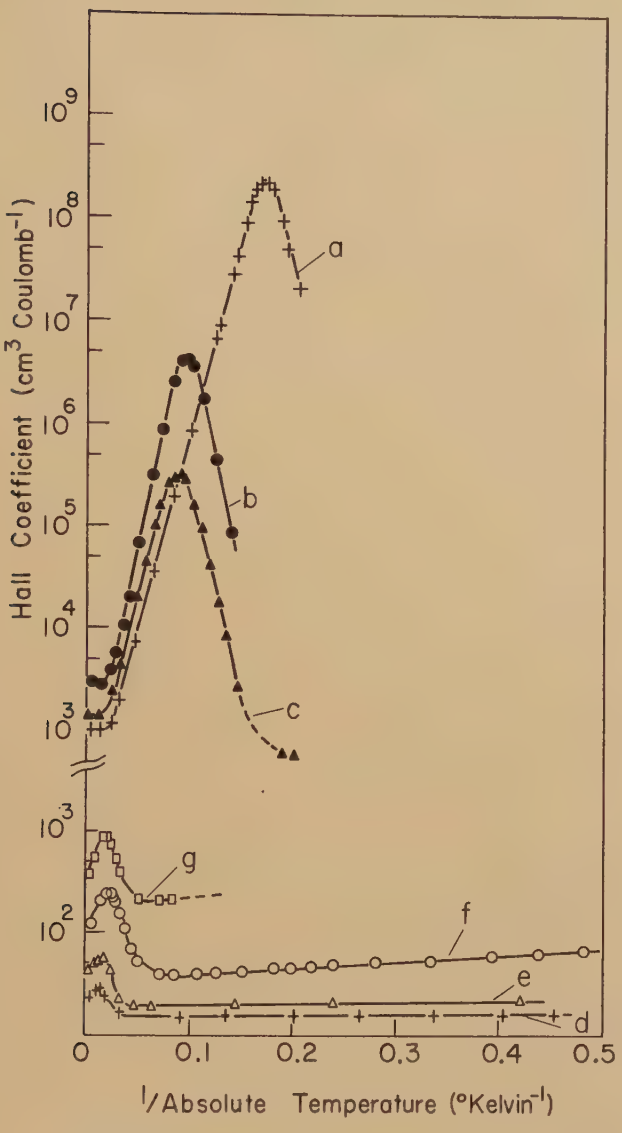
$$\rho = 1/ne\mu,$$

Fig. 9



Resistivity of p-type germanium with variable degree of compensation (Fritzsche and Lark-Horovitz 1959).

Fig. 10



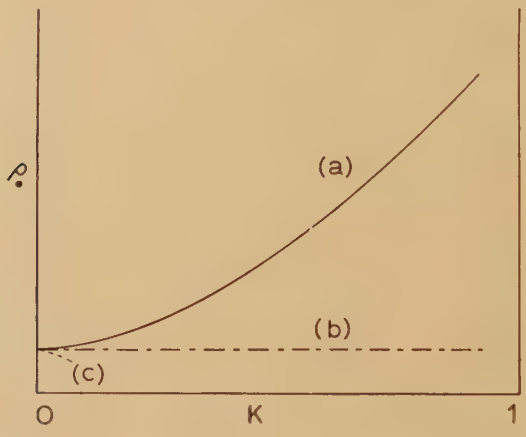
The Hall constant of the samples of fig. 9.

if the conduction is metallic ρ should always increase with K, assuming the carrier mobility μ to depend only weakly on K. If the conduction is non-metallic, ρ should decrease as K increases to $\frac{1}{2}$, and increase as K goes from $\frac{1}{2}$ to 1.

The dependence of ρ on K in the non-metallic range has not however been adequately explained. Since the term $\exp(\epsilon_3/kT)$ dominates ρ at low temperatures, and ϵ_3 is observed to have a minimum value for $K \simeq 0{\cdot}4$

(rather than $K = 0{\cdot}5$), a minimum is also observed in ρ at $K = 0{\cdot}4$. Although Miller and Abrahams (1960) have been able to calculate the observed dependence of ϵ_3 on K (fig. 8), in their theory the pre-exponential factor ρ_0 is independent of K. This independence is observed only for $K < 10^{-2}$, and then for specimens in the transition range, to which their theory does not apply. On the basis of the trapping model (§ 2), $\rho_0(K)$ is again predicted incorrectly. For in this model the number of free carriers is proportional to $K^{1/2}$ (eqn.(2)). When K is very small we can assume that the mobility of free carriers between sites at large distances from compensating centres is independent of K. Hence ρ_0 should be proportional to $K^{-1/2}$, whereas ρ_0 is observed to increase with K (fig. 11).

Fig. 11



Variation of the pre-exponential factor ρ_0 (in the equation $\rho = \rho_0 \exp(-\epsilon_3/kT)$) with compensation K; (*a*) experimental variation, (*b*) calculated by Miller and Abrahams (1960) and (*c*) calculated on the trap model.

We have seen that the magnitude of the impurity conduction depends sensitively on the impurity concentration and degree of compensation. A further variable is the type of impurity. For example, when corrections to the effective-mass formalism (§ 3) are taken into account, it is found that a donor electron is localized in a smaller volume around an arsenic impurity atom than around antimony. Hence the resistivity is larger for arsenic than for antimony impurities, if the same number of both are present, due to the smaller overlap of the arsenic wave functions with neighbouring states.

The Bohr radius of the impurity wave functions in silicon is of order $\frac{1}{3}$ that in germanium. Hence the impurity conductivity in silicon is always very much smaller than in germanium, for comparable impurity concentrations. Also, the energy ϵ_1 required to activate electrons into the conduction band is of order 5×10^{-2} ev in silicon and 1×10^{-2} ev in germanium. Therefore the onset of impurity conduction is observed at higher temperatures

in silicon than in germanium (again for comparable impurity concentrations) due the faster freeze-out of electrons from the conduction band of silicon.

The impurity concentration at which the transition from non-metallic to metallic conduction is observed depends on the overlap of neighbouring donor states, and on the degree of compensation. Because the impurity states are more localized in silicon than in germanium, the transition is observed at higher impurity concentrations in silicon. Similarly, because the impurity Bohr radius is different for different impurities, the transition concentration is also a function of the type of impurity. Fritzsche and Cuevas (1960 b) finds that the activation energy ϵ_3 in p-type (gallium doped) germanium disappears at $N_A = 1{\cdot}09 \times 10^{17}\,\mathrm{cm}^{-3}$ when $K \simeq 0{\cdot}04$, and at $N_A = 1{\cdot}80 \times 10^{17}\,\mathrm{cm}^{-3}$ when $K = 0{\cdot}4$. When $N_A = 2{\cdot}5 \times 10^{17}\,\mathrm{cm}^{-3}$, the activation energy disappears at a value of K in the range 0·4 to 0·7 (Fritzsche and Lark-Horovitz 1959). These results are discussed in more detail in § 11.

It is found that there is a large change in impurity resistivity when the overlap of neighbouring impurity states is altered by straining the crystal. We saw in § 3 that the wave function ψ of an isolated donor can be constructed from Bloch orbitals taken from the degenerate conduction band minima. Although the envelope functions F (eqn. (8)) are anisotropic, in an unstrained crystal ψ has the tetragonal symmetry of the lattice, and the overlap of neighbouring donor states leads to an isotopic resistivity. In a germanium crystal strained along, for example, the [110] direction, two conduction band minima are depressed in energy relative to the other two. If the strain is large enough, the donor electron ground state will be a sum of the $F_i \phi_i$ corresponding to the depressed minima only,

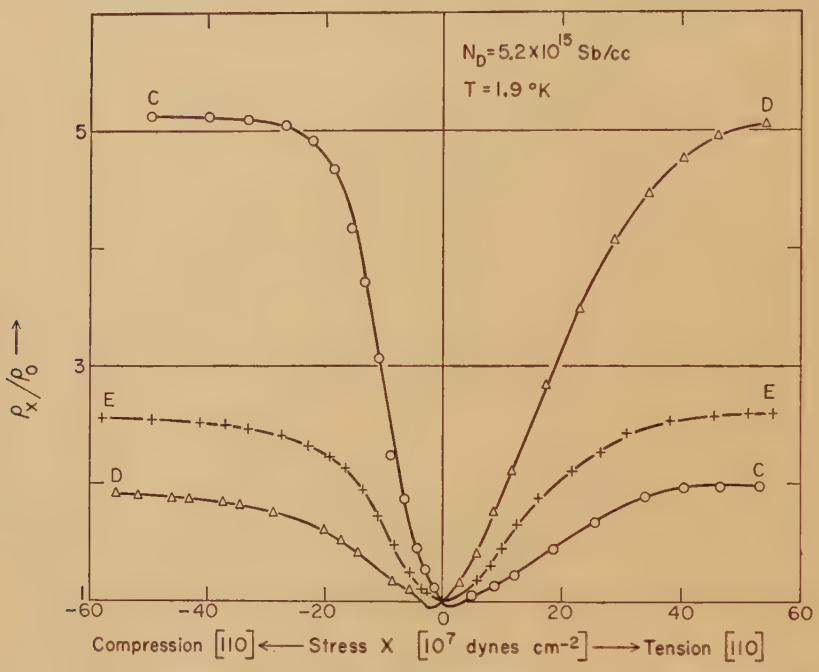
$$\psi_S = 2^{-1/2}(F_i \phi_i + F_j \phi_j).$$

The overlap of such functions will vary strongly with the direction to the neighbouring donor relative to the strain direction. Hence we expect that in a strained single crystal the impurity resistivity will be anisotropic, with a maximum value in the direction of least overlap. Fritzsche (1960 b) has measured the change in impurity conduction in germanium containing $5{\cdot}2 \times 10^{15}$ antimony atoms per unit volume at 1·9°K as a function of uniaxial tension and compression along the [110] direction. Figure 12 shows the resistivity ratios of the strained to the unstrained sample for the three principle directions of the resistivity tensor (labelled C, D and E). A 'saturation' is observed when strains are large enough for only the depressed minima to contribute to the donor wave function. The relative magnitudes of ρ_C, ρ_D and ρ_E are in agreement with predictions on the basis of the overlap of these functions.

Additional evidence that there is a conduction process in the impurity energy levels at low temperatures is provided by the work of Sladek (1956, 1958, 1959) on magnetically induced impurity banding in n-type indium antimonide. Because of the small effective mass ratio ($m^*/m = 0{\cdot}013$) and

large dielectric constant ($\kappa = 16$), the Bohr radius of a donor electron in indium antimonide is large, about 144 Å, and the ionisation energy of an isolated donor is small (0·0067 ev). For all obtainable purities the overlap of donor wave funtions is large and the levels are broadened and merge with the bottom of the conduction band. A strong magnetic field will shrink the donor wave funtions (Yafet *et al.* 1956), producing two effects.

Fig. 12

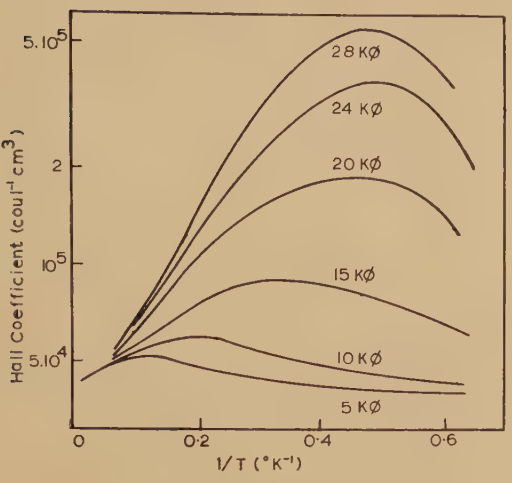


Ratio of the resistivity of stressed and unstressed n-type germanium, for the principle directions, as a function of stress (Fritzsche 1960 b).

(i) The donor ionization energy is increased due to the decrease in the Coulomb energy of the electron. Thus for a large enough magnetic field donor levels will be split off from the bottom of the conduction band.

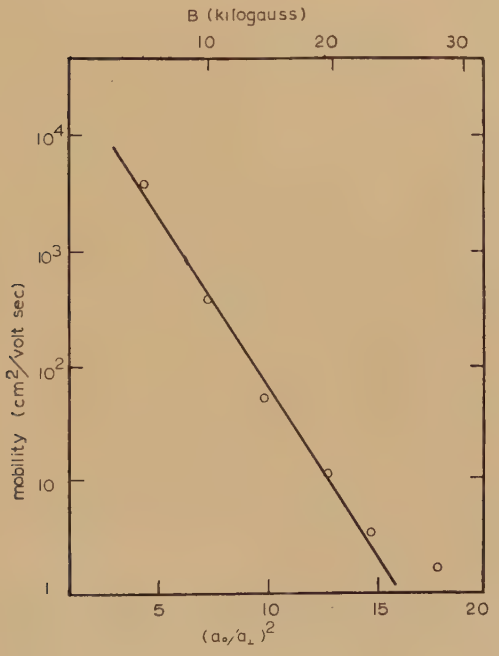
(ii) The overlap of the donor wave functions is decreased. This effect is large because of the small effective mass. Figure 13 shows measurements of the Hall coefficient R as a function of temperature and magnetic field strength, in a specimen containing $5{\cdot}3 \times 10^{14}$ donors per unit volume (Sladek 1958). For a small magnetic field R is approximately temperature-independent since the specimen is degenerate. At fields greater than about 10 k-gauss we see the onset of typical impurity conduction. R reaches a maximum value at around 20°K due to a competing conduction process in the split-off donor levels. The depression of the donor levels below the conduction band increases with the magnetic field B, as shown by the increasing slope of the Hall curve with B at temperatures above the Hall

Fig. 13



The Hall constant as a function of magnetic field in InSb containing $5{\cdot}3 \times 10^{14}$ donors/cm^3 (Sladek 1958).

Fig. 14



The effect of a magnetic field B transverse to the current direction on the mobility of impurity electrons in n-type InSb; $a_\perp$ is the Bohr radius of a donor wave function in a direction perpendicular to B (Sladek 1958).

maximum. The Hall mobility μ_H for the same specimen measured at a temperature below the Hall peak, is plotted in fig. 14 against $(a_0/a_\perp)^2$. The Bohr radius a_0 is the zero field value, while $a_\perp$ is the radius normal to the field H, calculated from the theory of Yafet *et al.* (1956). Thus as the magnetic field shrinks the donor electron's orbit and hence lowers the overlap of neighbouring donor states, the electron's mobility decreases, as is to be expected on the impurity conduction model.

§ 5. Methods of Calculating the Electrical Conductivity at Low Concentrations

By low concentrations we mean here concentrations such that the electron gas is not 'metallic', so that the conductivity tends to zero with the temperature. In this range of concentration we have to do essentially with a 'one-body' problem, the movement of a single electron under circumstances in which the interaction between electrons is not important (except in so far as an occupied centre blocks the passage of an electron from another centre). However, even so, the problem is complicated, and far from a complete solution.

If we take as our first problem the movement of an electron in a disordered lattice in which the ions are assumed to be at rest, we have to ask whether the states are *localized* or unbounded in space. By disordered, we mean either in random positions or acted on by a random field (such as that from the charged minority centres) or both. By a 'localized' state we mean that each characteristic solution of the Schrodinger equation for an electron in this field decays exponentially to zero at sufficiently large distances from some point in space. The problem of the conditions under which states are localized or not is by no means solved. In § 7 we reach the conclusion that *all* states in a one-dimensional lattice may be localized. This may correspond to the theorem that in one dimension any potential hole, however small, leads to a bound state. In three dimensions the work of Anderson (1958) was the first to show that, if the separation between impurity centres is large enough, all states would be localized, but at some higher concentration one would go over to unbound states†. Anderson was concerned with spin diffusion; Twose (1959) extended his work to impurity conduction, finding as the criterion for bound states

$$N_t < 10^{-4}/a_0^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where a_0 is the hydrogen radius of the centres $(\hbar^2\kappa/m^*e^2)$. Due to the approximations that it was necessary to make in this work, the result may be too small‡.

Suppose now that (9) is not satisfied, so that an electron has a wave function extending through the lattice. We shall now consider the

† This concentration has nothing to do with the concentration for metallic conduction discussed in part II.

‡ See Appendix for detailed discussion.

problem of its mobility, under conditions when the interaction with other electrons is not important (i.e. $1-K \ll 1$, or for a 'metallic' electron gas).

We shall have to calculate the time of relaxation τ and mean free path l. The question arises as to whether the usual transport theory can be used, developed as it is for the case where the scattering processes are weak compared with those interactions between lattice atoms which broaden the sharp electronic levels of the isolated atom into a band in the solid. If τ_s is an average relaxation time for a carrier scattered between states of different energy and momentum in the band, a suitable criterion that the band approach be a good one is (Joffe 1956, Herring 1959, Yamashita and Kurosawa 1960)

$$\tau_s \gtrsim \hbar/W,$$

otherwise the uncertainty in energy of the scattered carrier would exceed the bandwidth. This leads to a lower limit on the carrier mobility μ below which the band model will not be an adequate starting point for a conductivity calculation. For

$$\mu = \frac{e\tau_s}{m^*} \gtrsim \frac{e\hbar}{m^*W} \sim \frac{ea^2}{6\hbar}, \quad \ldots\ldots (10)$$

where m^* is the effective mass of the carrier and a the lattice spacing. Here we have used the Bloch tight-binding model to estimate m^*W; this should be a valid approximation in those substances for which the band approach is questionable, for example, the narrow 3d 'band' in transition metal oxides. With $a \sim 4$ Å, the mobility must be larger than of order

$$\mu_L = 0{\cdot}5 \text{ cm}^2/\text{volt sec}. \quad \ldots\ldots (11)$$

Similarly, the carrier mean free path $l \sim \tau_s \bar{v}$ (where $\bar{v}$ is the mean velocity) must be larger than the order of the lattice spacing:

$$\left.\begin{aligned} &\bar{v} \sim \pi\hbar/m^*a, \\ &l \sim \frac{\pi\hbar\tau_s}{m^*a} \gtrsim \frac{\pi\hbar^2}{am^*W} \sim \frac{\pi a}{6}. \end{aligned}\right\} \quad \ldots\ldots (12)$$

The numerical factors are of course approximate. It is possible to define the wave vector k of a Bloch electron only to within $1/l$, on a similar argument.

In metals, these conditions are usually easily satisfied. Electrons in copper, for example, have a mean free path of order 140 lattice spacings and mobilities 30 cm^2/volt sec at room temperature. In a large number of semiconductors however, the carrier mobility may be of order μ_L or lower. Taking the transition metal oxides as examples, in $L_{0.1}i\,Ni_{0.9}\,O$ carriers in the 3d (Ni) levels have mobilities ranging from 10^{-3} to 7×10^{-2} cm^2/volt sec from room temperature to 1000°K (Morin 1958). In titanium oxide (TiO) Morin estimates the 'd band' mobility as 0·4 cm^2/volt sec at room temperature. In the case of impurity conduction in germanium or silicon, the condition (10) for the mobility becomes

$$\mu_L \gtrsim \frac{e}{6\hbar}\left(\frac{3}{4\pi N}\right)^{2/3} = 10^{14} N^{-2/3} \text{ cm}^2/\text{volt cm}. \quad \ldots (13)$$

Here we have assumed that the majority impurities, of concentration N, lie on a lattice of average spacing $(3/4\pi N)^{1/3}$. Table 2 compares experimental estimates of the mobility, μ_{exp}, with μ_L for n-type germanium. Thus a treatment of impurity conduction which assumes that the impurity electrons can be described by modified Bloch-type functions would appear to break down as N decreases into the transition region.

Table 2

N (cm^{-3})	10^{15}	10^{16}	10^{17}	10^{18}
μ_L (cm^2/volt sec)	10^4	2×10^3	5×10^2	10^2
μ_{exp} (cm^2/volt sec)	5×10^{-6}	1·0	$4\cdot10^2$	10^3

Methods based on the density matrix have been developed to handle the conductivity of solids under these conditions; we have thought it worth while to include in the next section a simplified treatment, limited to one dimension, to show what these methods mean.

We turn now to the case when the states are localized. Then in general a finite amount of energy is necessary to move an electron from one localized state to the next. This can only come from phonons. We have thus to consider the interaction with phonons.

Phonons can act in two ways, which can be distinguished in the language of field theory as strong and weak interactions. In polar lattices it has been known for a long time that 'self-trapping' is a possibility (Landau 1933, Mott and Gurney 1940, p. 86). The electron or positive carrier polarises the lattice round it and can only move by carrying this polarization with it. A jump from one site to another is thus a multiphonon transition. This is believed to occur in nickel oxide, the carrier being a positive vacancy on a Ni^{2+} ion.

Single phonon transitions are probably of predominating importance in valence semiconductors, and these are discussed in § 7. We start here with the concept of a localized state; an electron is localized in a given centre and, receiving energy from a phonon, it makes a transition by tunnel effect to another centre where the energy is different. This is common to the treatment of Twose and of Miller and Abrahams: the latter are interested in the case where K is not small and so all centres are not available for any one moving electron or hole.

§ 6. The Properties of a One-dimensional Disordered Lattice

In this section we investigate some of the properties of a one-dimensional lattice, not with a view to applying them to the actual problem, but as an illustration of some of the principles involved.

We shall first (in § 6.1) investigate the mobility of an electron in a one-dimensional lattice, assuming that the wave functions are extended

throughout the lattice. The purpose of the investigation is to give a method appropriate to the case when the mean free path is comparable with the electron's wavelength, so that the usual transport theory based on Boltzmann's equation is not applicable. The calculation for three dimensions has been given by Edwards (1958), but that in one dimension is so much simpler that it is worth reproducing.

In § 6.2 we shall examine the nature of the wave functions in a one-dimensional disordered lattice. There have been many investigations of the density of states in this case (for refs see Frisch and Lloyd 1960) but as far as we know none of the nature of the states. We shall show that in many cases (perhaps in all) the states are bound.

6.1. *The Conductivity of Electrons in a One-dimensional Disordered Lattice*

Throughout this section we discuss the conductivity of carriers on a one-dimensional 'wire' of length L, on which they have a non-periodic potential energy $V(x)$. We have to develop a transport theory appropriate to this case. Before doing this we shall set down the usual transport theory based on Boltzmann's equation in a form appropriate to a one-dimensional model.

In this case, where the state of the carrier can be defined by a wave number k and a time of relaxation τ can be defined, the current j is given by

$$j = e \int_{-\infty}^{\infty} N(k) f(k)\, u\, dk \qquad \ldots\ldots (14)$$

where $N(k)\,dk$ is the number of states in the range k to $k+dk$, $f(k)$ the probability that a state k is occupied and u the velocity ($u = dE/\hbar dk$). If a field F is applied, then in a steady state

$$f = f_0 + (df/dk) eF\tau/\hbar,$$

where f_0 is the form of f in the absence of a field. Thus from (14) we obtain

$$j = \frac{e^2F}{\hbar} \int N(k) \frac{df}{dk} u\tau\, dk. \qquad \ldots\ldots (15)$$

We may write $u\tau = l$, where l is the mean free path, and

$$N(k) = L,$$

so that

$$j = \frac{e^2FL}{\hbar} \int l \frac{df}{dE} dE. \qquad \ldots\ldots (16)$$

This formula gives the total current in the wire. The current C at any point is obtained by dividing by L, so that

$$C = \frac{e^2F}{\hbar} \int l \frac{df}{dE} dE. \qquad \ldots\ldots (17)$$

If we wish to derive a mobility from these formulae, we may suppose a Boltzmann distribution of electrons and set

$$f = \text{const.} \exp(-E/kT),$$

so that, from (16)

$$j = \frac{e^2F}{h}\int N(k)\frac{df}{dE}l\frac{dE}{dk}\,dk = e^2F\int N(k)flu/kT\,.\,dk,$$

from which we see that the mobility μ is given by

$$\mu = elu/kT \sim el/m^*u \sim e\tau/m^*,$$

where m^* is the effective mass.

We shall now derive by a simplified method a formula due to Greenwood (1958) for the current C, valid even when a mean free path cannot be defined. We assume as before that the electron moves on a circular wire of length L on which its potential energy $V(x)$ is that of some random (non-periodic) field. Then the Schrödinger equation of such an electron is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V)\psi = 0.$$

The solutions are subject to the boundary conditions

$$\psi(x) = \psi(x+L), \quad \psi'(x) = \psi'(x+L),$$

so that they are single-valued on the circular wire; they are real and in general non-degenerate. We denote the stationary wave functions and energy states by ψ_n, E_n. As in Greenwood's paper we imagine an electromotive force induced in the circuit by a magnetic field, threading the circuit and increasing uniformly with the time. The perturbing term in a time-dependent Hamiltonian is

$$(\hbar eF/mi)\,t\,\partial/\partial x,$$

where F is the induced electric field and t the time. If then an electron is initially in the state n, at a time t after the perturbation is switched on, the wave function Ψ is given by

$$\Psi = \psi_n \exp(-iE_nt/\hbar) + \sum_{n'} a_{nn'}(t)\,\psi_{n'}(x)\exp(-iE_{n'}t/\hbar),$$

where

$$(\hbar/i)\,a_{nn'}(t) = (\hbar eF/mi)D_{nn'}\int_0^t t\exp(i\omega_{nn'}t)\,dt.$$

Here

$$D_{nn'} = \int_0^L \psi_{n'}\frac{\partial}{\partial x}\psi_n\,dx,$$

and

$$\omega_{nn'} = (E_{n'} - E_n)/\hbar.$$

Integration by parts and neglect of oscillating terms as in Greenwood's paper gives

$$a_{n'} = (eF/m)\,D_{nn'}\{1 - \exp(i\omega_{nn'}t)\}/\omega^2{}_{nn'}.$$

The current due to an electron initially in state n is

$$\frac{e\hbar}{2mi}\int\{\Psi^*\frac{\partial}{\partial x}\Psi - \Psi\frac{\partial}{\partial x}\Psi^*\}\,dx = (e^2\hbar F/m^2)\sum|D_{nn'}|^2\sin\omega_{nn'}t/\omega^2{}_{nn'}.$$

If f_n is the occupation number of each state, the resultant current may be written

$$\frac{e^2\hbar^3F}{m^2}\sum_n\sum_{n'}|D_{nn'}|^2\frac{f_n-f_{n'}}{E_n-E_{n'}}\frac{\sin\{(E_n-E_{n'})t/\hbar\}}{E_n-E_{n'}}. \quad . \quad . \quad (18)$$

If $N(E)$ is the density of states, then for large values of the time t this becomes

$$\frac{\pi e^2\hbar^3F}{m^2}\int\overline{|D_{nn'}|^2}\frac{\partial f}{\partial E_n}\{N(E_n)\}^2\,dE_n, \quad . \quad . \quad . \quad . \quad (19)$$

where the bar gives an average over values of n' near to n. This formula is valid in general; it will be noted that m is the electronic mass, not the effective mass. Also to obtain the current C at any point we must divide by L.

Formula (19) is the basic expression for conductivity in one dimension. With any such formula, our first task is to show that for small perturbing energy V it leads to the formula given by the Boltzmann equation. This has been done, in the general three-dimensional case, by Edwards (1958). His proof uses advanced methods, and we shall now give a simple discussion to show how the Boltzmann expression (17) arises.

For any form of $V(x)$ we can define a wave number k such that $1/k$ is the mean distance between zeros of ψ_n. Then as before

$$N(E)=L/(dE/dk).$$

If the disordered field is small perturbation on a periodic field in which the electron has an effective mass m^*

$$N(E)=Lm^*/\hbar^2k. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

We shall now estimate $|D_{nn'}|^2$ for this case. First we note that, while the diagonal element D_{nn} vanishes, the off-diagonal element $D_{nn'}$ is not small even if n and n' differ only by unity, since the phase of ψ_n may differ from that of $\psi_{n'}$ by a large amount. Thus the integral

$$\int\psi_n(\partial\psi_{n'}/\partial x)\,dx,$$

integrated over one mean free path l, a distance in which the wave functions are coherent, will be of order

$$lmk/m^*L; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

the quantity L in the denominator comes from the normalizing factor of the functions ψ_n. There are L/l mean free paths in the length in the wire, and the signs of the contribution of each to D will be random; thus, to obtain the root mean square of D, (21) should be multiplied by $(L/l)^{1/2}$, and

$$|D_{nn'}|\sim(l/L)^{1/2}mk/m^*. \quad . \quad . \quad . \quad . \quad (22)$$

Substituting for $|D_{nn'}|^2$ and for $\{N(E)\}^2$ in eqn. (19) and dividing as before by L to get the curent at any point, we find

$$C=\frac{e^2F}{\hbar}\int l\frac{df}{dE}\,dE,$$

apart from numerical factors, which is identical with (17).

It is of interest to show directly that the mean free path as usually defined can be equated (apart from a numerical constant) with the distance in which any given phase relationship between two functions, ψ_n, $\psi_{n'}$ of very nearly the same energy is lost. For one dimension we shall treat the problem by considering a row of scattering centres at points $x_1, x_2, \ldots x_n$, each reflecting a *small* proportion of any incident wave. It is convenient to use at each of these centres the delta-function used by Lax and Phillips (1958) so that at each centre ψ is continuous and ψ' changes by $\Delta\psi$, where

$$\Delta\psi'/\psi = \eta$$

and η is small.

Suppose that such a centre is situated at $x = 0$ and we ask for the amplitude of the wave reflected from it. We set for an incident and reflected wave $(x < 0)$

$$\psi = \exp(ikx) + A\exp(-ikx)$$

and for a transmitted wave $(x > 0)$

$$\psi = B\exp(ikx).$$

The boundary conditions give

$$1 + A = B, \qquad 1 - A = B + q$$

where $q = \eta/k$. Hence

$$A = \tfrac{1}{2}q.$$

Thus the amplitude reflected by the first N of such centres is

$$\tfrac{1}{2}q\sum_{1}^{N}\exp(2ikx_N).$$

An estimate of the number N of centres in one mean free path would be obtained by equating this to unity.

Now consider the phase of a *real* function ψ and let us ask how much it changes at a scattering centre. Let us write for $x < 0$

$$\psi = A_1\cos(kx + \zeta_1)$$

and for $x > 0$

$$\psi = A_2\cos(kx + \zeta_2).$$

The boundary conditions give

$$A_1\cos\zeta_1 = A_2\cos\zeta_2, \quad \tan\zeta_1 - \tan\zeta_2 = \eta/k = q,$$

and hence, since q is small

$$\Delta\zeta = \zeta_1 - \zeta_2 = q\cos\zeta_1\cos\zeta_2.$$

Since ζ_1 and ζ_2 are nearly equal, this may be written

$$\Delta\zeta = \tfrac{1}{2}q(\cos 2\zeta - 1).$$

The second term (-1) in the bracket obviously makes no contribution to the rate at which two waves get out of phase with each other. The total shift in the phase of two waves initially $\frac{1}{2}\pi$ out of phase with each other will be the difference between $\frac{1}{2}q\cos 2kx_n$ and $\frac{1}{2}q\sin 2kx_n$, which will be of order unity in one mean free path as defined above. Either method gives

for the mean free path a quantity of order

$$l \sim b^3 q^2/(\Delta b)^2,$$

where b is the mean distance between centres, Δb the root mean square fluctuation and $q = \eta/k$ as before.

A particularly interesting application of these ideas is to the 'tight binding' case, which of course is applicable to impurity-band conduction. The overlap integral between two hydrogen-like centres, of Bohr radius a, distance b apart, is

$$j = 2W_0(1 + b/a)\exp(-b/a).$$

Elementary considerations based on a deformation potential suggest that, for a mean displacement Δb from positions in a periodic lattice, the mean free path will be of order

$$b \Big/ \left(\frac{\Delta j}{j}\right)^2, \quad \Delta j = \frac{\partial j}{\partial b}\Delta b.$$

The mean free path l will be of order b when $\Delta j \sim j$ and if $b \gg a$ this will occur when $\Delta b \sim a$, and thus for $\Delta b/b \ll 1$. It is possible—and the experimental evidence reviewed in § 12 suggests that this is the case—that l may become less than b before bound states occur—that is to say, in the three-dimensional case. If so, we shall want to know what happens to formula (22) in this case. No detailed calculations have been given, but we should expect a random fluctuation of the *amplitude* of the wave function on each centre by $\exp(-\Delta b/a)$. Since the normalization of the wave function will be determined by the larger amplitudes, typical terms in the integral for $D_{nn'}$ will be reduced by just this factor. We thus expect the mean free path l to be of order $l \sim b\exp(-2\Delta b/a)$.

6.2. *Bound States in the One-dimensional Model*

The above analysis assumes that the wave functions extend through the lattice (the states are not bound). Actually in one dimension this may not be the case, but the analysis is of some interest on account of its possible extension to three dimensions.

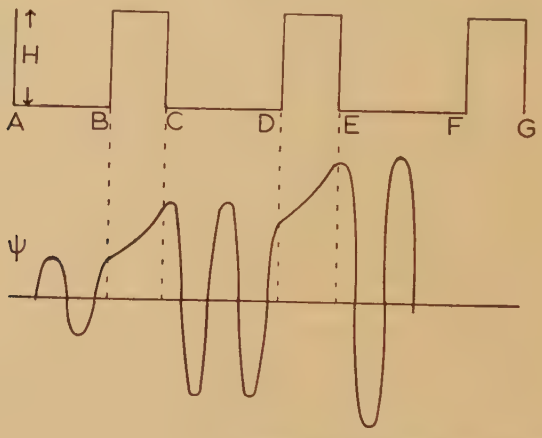
The simplest model for which this may be seen is that in which we take for V a potential of Kronig–Penney type, as shown in fig. 15, namely a series of potential barriers, a mean distance b apart, but such that the root mean square of b is Δb. We supposed that Δb is greater than or comparable with the wave length of ψ in the range AB. We suppose that the two solutions, which increase or decrease exponentially in the range BC, increase or decrease by p or $1/p$, and that p is large.

Consider an oscillating solution in the range AB of form $\psi = \cos(kx + \zeta)$. Then for all phases $\zeta\ (-\pi \leqslant \zeta \leqslant \pi)$, except those in a range of order $1/p$, ψ will increase in the range BC. If ψ has increased in the range BC, then, whatever the phase in AB, ψ will in general increase again by the factor p in DE. Thus the solution with arbitrary phase in AB increases exponentially as x increases, and by the same argument will increase exponentially as x diminishes. Such a solution is shown in fig. 15 and obviously has no physical significance.

As we have seen, for all phases ζ in AB except those in a range $\Delta\zeta \sim 1/p$, ψ increases in BC; and for these solutions ζ varies only by $1/p$ in CD, by $1/p^2$ in EF and so on. Thus our typical solution, obtained by starting with arbitrary energy and phase in an interval such as AB, will increase exponentially in both directions, the phase becoming more closely defined as the distance from AB increases. It is of course possible to choose the phase in AB so as to give an exponential solution which increases or decreases over any relevant range of x.

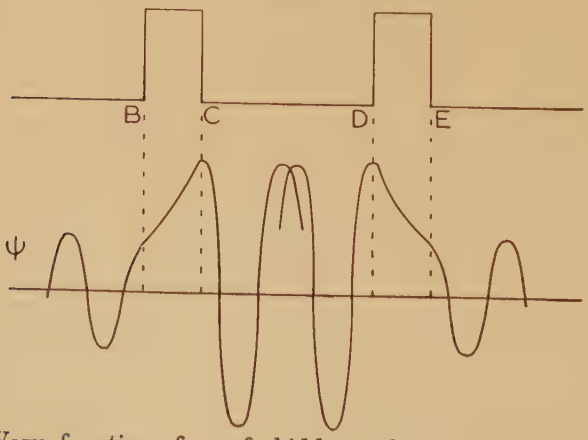
The stationary solution may be found as follows. Take any gap in the potential energy, for instance CD. For any energy W, one can set up solutions from both sides which increase as one comes towards CD. In general these solutions will not fit in the middle (fig. 16), but one can choose values of the energy such that they do. These are the quantized energy values of the problem.

Fig. 15



Wave function in a one-dimensional disordered lattice.

Fig. 16



Wave functions for a forbidden value of the energy.

Our conclusion that the stationary solutions of the problem correspond to bound states thus holds unless there is an accidental degeneracy. Although for a one-dimensional space of length L the energy levels are spaced very closely, such that

$$\Delta W \sim W b/L,$$

two localized wave functions distant in space na apart must have the same energy to any accuracy of order $W(1/p)^n$ to give a resonance, and this will be unlikely.

We have assumed that $k\Delta b \sim 1$. If $k\Delta b$ is small and the energy is chosen in the allowed band-width, it is clear that the average increase of ψ in going through each barrier is not

$$p = \exp\left[b\{2m(H-E)/\hbar^2\}^{1/2}\right],$$

where H is the height of a barrier, but of order

$$\exp\left[\Delta b\{2m(H-E)/\hbar^2\}\right].$$

For a periodic field ($\Delta b = 0$) the states are of course not localized. We have not been able to extend the proof to a more general case.

§ 7. The Interaction of Localized Carriers with Lattice Vibrations

In this section we shall be dealing throughout with a three-dimensional lattice. Our problem is to discuss the movement of an electron from centre to centre under conditions in which the states are localized, so that lattice vibrations are essential in making the transitions possible. We have to distinguish between strong coupling, appropriate to ionic lattices, and weak coupling, appropriate to silicon and germanium. In the former case, we have to show how the concept of a polaron arises.

In the strong coupling case we may consider the compound $Li_x Ni_{1-x} O$. Here a small fraction of Ni^{2+} ions have been replaced by Li^+ ions, with the formation of Ni^{3+} ions to preserve electrical neutrality. Each of these can be regarded as a positive hole bound to a Ni^{2+} site. We discuss the motion of this hole through the lattice of Ni^{2+} ions; the potential that hole experiences is not perfectly periodic, due to the random arrangement of Li^+ ions. This problem has been treated by Yamashita and Kurosawa (1958, 1960) and Sewell (1958). Since the carrier-lattice coupling is strong it cannot be treated as a perturbation, as is done in the usual theory of conduction by Bloch electrons. In a more exact treatment of the lattice coupling, we find that the hole bound to a given Ni^{2+} ion moves in the potential of that ion and of the polarization field produced by the displacement of the surrounding ions from their equilibrium positions. Thermal vibrations of the lattice ions are about the displaced positions. Thus the bound hole can be represented by the wave function

$$\phi_i(\mathbf{r}-\mathbf{R}_i)\prod_\sigma X_{n_\sigma}\{Q_\sigma(\mathbf{R}_i)\}, \quad \ldots\ldots \quad (23)$$

which includes explicitly the interaction between lattice and hole. Here $\mathbf{R}_i$ is the position of the ith nickel ion, $\mathbf{r}$ the coordinate of the hole, Q_σ a phonon coordinate and n_σ the number of phonons in the mode σ. The Hamiltonian of the hole can be written

$$\mathscr{H} = \mathscr{H}_{\mathrm{h}} + \mathscr{H}_{\mathrm{int}} + \mathscr{H}_{\mathrm{L}},$$

where

$$\mathscr{H}_{\mathrm{L}} = \tfrac{1}{2}\sum (P_\sigma^2 + \omega_\sigma^2 Q_\sigma^2),$$

so that $\mathscr{H}_{\mathrm{L}}$ describes the lattice vibrations and Q_σ is the displacement of the vibration σ. The term

$$\mathscr{H}_{\mathrm{h}} = p^2/2m + \sum_j U(\mathbf{r}-\mathbf{R}_j) + \Delta U(\mathbf{r}),$$

describes the hole, with momentum p, moving in the field of the potential energy $U(\mathbf{r}-\mathbf{R}_j)$ due to nickel ions at points $\mathbf{R}_j$. The sum is over all lattice sites j. ΔU describes the corrections to the periodic potential due to other positive holes and Li^+ ions. The lattice-hole interaction is of the form

$$\mathscr{H}_{\mathrm{int}} = \sum C_\sigma \exp(i\boldsymbol{\tau}_\sigma \mathbf{r})\, Q_\sigma, \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where $\boldsymbol{\tau}_\sigma$ is the wave vector of mode σ, and C_σ includes the coupling constant. A variational technique (see, for example, Fröhlich 1954) is used to find an eigenfunction in the form $\phi_i(\mathbf{r}-\mathbf{R}_i)X_n(Q^{(i)})$ to the Hamiltonian

$$\mathscr{H}^{(i)} = \mathscr{H}_{\mathrm{h}}^{(i)} + \mathscr{H}_{\mathrm{int}} + \mathscr{H}_{\mathrm{L}},$$

where

$$\mathscr{H}_{\mathrm{h}}^{(i)} = p^2/2m + U(\mathbf{r}-\mathbf{R}_i) + \Delta U(\mathbf{r}).$$

Then ϕ_i is required to be a solution of

$$\{p^2/2m + U(\mathbf{r}-\mathbf{R}_i) + \Delta U(\mathbf{r}) + \langle X^* \,|\, \mathscr{H}_{\mathrm{int}} \,|\, X\rangle\}\phi(\mathbf{r}-\mathbf{R}_i) = \epsilon_i \phi_i(\mathbf{r}-\mathbf{R}_i). \quad (25)$$

The term $\langle X^* \,|\, \mathscr{H} \,|\, X\rangle$ describes the polarization field of the surrounding ions. We see that ϕ_i is predominantly just the wave function of a hole localised in the potential $U(\mathbf{r}-\mathbf{R}_i)$. Similarly, $X_n(Q^{(i)})$ must satisfy

$$\{\tfrac{1}{2}\sum(P_\sigma^2 + \omega_\sigma^2 Q_\sigma^2) + \langle \phi_i^*(\mathbf{r}) \,|\, \mathscr{H}_{\mathrm{int}} \,|\, \phi_i(\mathbf{r})\rangle\} X_n = \epsilon_n X_n. \quad . \quad . \quad (26)$$

From (24),

$$\langle \phi_i^* \,|\, \mathscr{H}_{\mathrm{int}} \,|\, \phi_i\rangle = \sum C_\sigma \langle \phi_i^* \,|\, \exp(i\boldsymbol{\tau}_\sigma \mathbf{r}) \,|\, \phi_i\rangle\, Q_\sigma = \sum_\sigma A_\sigma^{(i)} Q_\sigma. \quad . \quad (27)$$

By a change of coordinates, namely

$$Q_\sigma = Q_\sigma^{(i)} - C_\sigma^{(i)}, \quad C_\sigma^{(i)} = A_\sigma^{(i)}/\omega_\sigma^2, \quad . \quad . \quad . \quad . \quad (28)$$

(26) becomes

$$\tfrac{1}{2}\sum (P_\sigma^2 + \omega_\sigma^2 Q_\sigma^{(i)2})\, X_n^{(i)} = \epsilon_n'\, X_n^{(i)}.$$

Thus X_n is a product of single mode oscillator functions,

$$X_n^{(i)} = \prod X_{n_\sigma}[Q_\sigma^{(i)}],$$

each having the displaced coordinate $Q_\sigma^{(i)}$. The energy corresponding to the state $\phi_i X_n$ is ϵ_{in}, where

$$\epsilon_{in} = \epsilon_i + \epsilon'_n = \epsilon_i + \sum(n_\sigma + \tfrac{1}{2})\hbar\omega_\sigma - \sum |A_\sigma^{(i)}|^2/2\omega_\sigma. \quad . \quad . \quad (29)$$

We can now treat the potential $(\mathscr{H}_h - \mathscr{H}_h^{(i)})$ due to the other lattice sites as the perturbation which causes the hole to move. This is analogous to the usual Bloch tight-binding approach. In the Bloch case, however, one starts with a product of an atomic orbital $\psi(r-R_j)$ and phonon functions $X_n(Q)$, X_n being independent of the site R_i. A hole (or electron) with wave vector k is the described by

$$\psi_k = \{\sum_i \exp(i\mathbf{k}\mathbf{R}_i)\,\psi(\mathbf{r}-\mathbf{R}_i)\}\,X_n(q)$$

and has energy, for example, for a simple cubic lattice with spacing a,

$$\epsilon_k = \epsilon_0 + 2M_B(\cos\mathbf{k}a).$$

Here

$$M_B = \{\phi_i(\mathbf{r}-\mathbf{R}_i)[\mathscr{H}-\mathscr{H}_h^{(i)}]\,\phi_j(\mathbf{r}-\mathbf{R}_i)\,d\tau\}\langle X_n \mid X_n\rangle, \quad . \quad (29a)$$

where sites i and j are nearest neighbours. We note that the bandwidth $12M_B$ and the effective mass m^* of the hole $(m^* = \hbar^2/(\partial^2\epsilon_k/\partial k^2)$ are independent of temperature, since $\langle X_n \mid X_n\rangle$ in the Bloch scheme. In a similar way, we can use the functions $\phi_i X_n^{(i)}$ to describe a polaron with wave vector k,

$$\psi_k = \sum \phi_i(\mathbf{r}-\mathbf{R}_i)\,X_n^{(i)}\exp(i\mathbf{k}\mathbf{R}_i),$$

and energy

$$\epsilon_k = \epsilon_0 + 2M\cos(\mathbf{k}a),$$

where now

$$M = M_B\langle X_n^{(i)} \mid X_n^{(j)}\rangle.$$

The polaron bandwidth and effective mass are temperature dependent, since the overlap $\langle X_n^{(i)} \mid X_n^{(j)}\rangle$ of the phonon functions is not equal to unity but is a function of the number n of phonons present. We may estimate the magnitude of this overlap at a temperature T by taking an average† value:

$$\langle X_n^{(i)} \mid X_n^{(j)}\rangle = \{\prod \langle X_{n_\sigma}(Q_\sigma^{(i)}) \mid X_{n_\sigma}(Q_\sigma^{(j)})\rangle\}_{\text{ave}}$$

$$= \prod_\sigma \{2\sinh(\hbar\omega_\sigma/2kT)\} \sum_{n_\sigma} \exp\{-(n_\sigma+\tfrac{1}{2})\hbar\omega_\sigma/kT\}\langle X_{n_\sigma}(Q^{(i)}) \mid X_{n_\sigma}(Q^{(j)})\rangle$$

$$= e^{-S}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where

$$S = \tfrac{1}{4}\sum (C_\sigma^{(i)} - C_\sigma^{(j)})^2\,\omega_\sigma(2\bar{n}_\sigma+1)/\hbar,$$

and $\bar{n}_\sigma$ is the mean number of phonons in mode σ at temperature T. The polaron effective mass m^* is therefore larger by a factor e^S than that calculated from the Bloch tight-binding model, and increases with temperature. This is because the moving hole must carry with it the polarisation of the surrounding lattice. Yamashita and Kurosawa (1958) have quoted values of S at $T=0$ (appropriate to $Li_xNi_{1-x}O$ ranging from 15 for the smallest value of x to 4–5 for the larger. Thus the mobility of a hole propagating as a wave may be very small.

† The averaging above can be performed by a method due to O'Rourke (1953).

The magnitude of S is a suitable criterion for giving meaning to the terms 'weak' or 'strong' coupling. The carrier-lattice coupling is strong if $S \gg 1$, and conversely when $S \ll 1$ the coupling is weak. This is a reasonable definition, for from (28, 30).

$$S \sim \tfrac{1}{2}\langle(\Delta l_\sigma/l_\sigma)^2\rangle,$$

where $\Delta l_\sigma \sim C_\sigma^{(i)}/M^{1/2}$ is the shift in the centre of oscillation of mode σ, $l_\sigma = (\hbar/M\omega_\sigma)^{1/2}$ is the mean amplitude of oscillation of mode σ at $T=0$ and the averaging is over all modes σ. Thus, for example, the polaron energy spectrum ϵ_k tends to that of an electron in the usual Bloch tight-binding approximation as $S \to 0$, as would be expected for small coupling.

In the following paragraphs we shall indicate how the hole mobility is calculated when the states are localized, so that the conduction mechanism is that of a hole jumping to neighbouring sites with the emission or absorption of phonons. The probability per unit time that a hole jumps from site i to j, with a change in the phonon state from n to n', is

$$W_{in,\,jn'} = (2\pi/\hbar)\,|M_{in,\,jn'}|^2\,\delta(\epsilon_{in} - \epsilon_{jn'}), \quad . \quad . \quad . \quad . \quad (31)$$

where the energies ϵ_{in} are given by (29), and

$$M_{in,\,jn'} = \langle\phi(\mathbf{r}-\mathbf{R}_i)X_n^{(i)}\,|\,\mathscr{H}_h - \mathscr{H}_h^{(i)}\,|\,\phi(\mathbf{r}-\mathbf{R}_j)X_{n'}^{(j)}\rangle.$$

To obtain the total transition rate W_{ij} of the hole from i to j, $W_{in,\,jn'}$ must be summed over final phonon states n' and averaged over n. Thus

$$W_{ij} = (2\pi/\hbar)\,|M_B|^2 \sum_{n,\,n'} p_n\,|\langle X_n^{(i)}\,|\,X_{n'}^{(j)}\rangle|^2\,\delta[\Delta\epsilon + \textstyle\sum(n_0 - n_0')\,\hbar\omega_0]. \quad (32)$$

Here M_B is as in (29 a), p_n is the probability that the state n is occupied at temperature T, and for the energy difference $\epsilon_{in} - \epsilon_{jn'}$ the value (29) has been substituted, and

$$\Delta\epsilon = \epsilon_i - \epsilon_j = \langle\phi_i\,|\,\Delta U(\mathbf{r})\,|\,\phi_i\rangle - \langle\phi_j\,|\,\Delta U(\mathbf{r})\,|\,\phi_j\rangle.$$

The evaluation of the summation in W_{ij} has been performed by several authors (references are given by Yamashita and Kurosawa 1960). We have

$$W_{ij} = \frac{1}{\hbar^2}\,|M_B|^2 \exp[-S(T)] \int_{-\infty}^{\infty} \exp[i\Delta\epsilon t/\hbar + G(T,t)]\,dt, \quad . \quad (33)$$

where

$$G(T,t) = \tfrac{1}{2}\textstyle\sum\{(2\bar{n}_\sigma + 1)\cos\omega_\sigma t + i\sin\omega_\sigma t\}\,\omega_\sigma (C_\sigma^{(i)} - C_\sigma^{(j)})(C^{(i)}C^{(j)})^2/\hbar \quad (34)$$

and $S(T)$ is as derived earlier (30). We note that an expansion

$$\exp[G(T,t)] = 1 + G + \tfrac{1}{2}G^2 + \ldots, \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

and a term by term integration over t gives essentially the contributions to W_{ij} due to zero, one, two . . . phonon transitions (since G is proportional, through $C_\sigma^{(i)}$, to the phonon-hole coupling constant squared). $G(T, t)$ can be seen to have a magnitude comparable to S for a range of t values around $t=0$. If t is outside this range, $G(T,t)$ tends rapidly to zero. As we have

seen, S can be as large as 15 for $Li_xNi_{1-x}O$. Hence if S is large enough the transitions in which phonons are emitted or absorbed will dominate the zero phonon rate, and as discussed in § 5 the electron will be 'self-trapped' at site i. Similarly, the higher terms in the expansion (35) (multiphonon transitions) will dominate the term linear in G (single phonon transitions).

The evaluation of W_{ij} from (33), which has been treated in the previous reference, is difficult and will not be discussed further here, since we are primarily interested in impurity conduction in valance semiconductors. We note that, assuming the Einstein relationship, a hole mobility μ may be calculated,

$$\mu = (ea^2/kT)W_{ij},$$

where a is the spacing of the Ni^{2+} ions (the mean jump distance of the holes).

We now discuss the case of *weak coupling*, appropriate to impurity conduction in valence semiconductors. We consider the motion of a hole bound to a donor atom (in n-type germanium with a small degree of compensation, say) surrounded by randomly placed charged donor and acceptor centres. By the same methods as in previous paragraphs we find that the strength of the interaction of the hole with the lattice depends on S, which can be estimated through the deformation potential constant E_1 (Bardeen and Shockley 1950). The 'displacement' $C_\sigma^{(i)}$ occurring in S is, from (25–28),

$$\begin{aligned} C_\sigma^{(i)} &= \frac{M^{1/2}\omega_\sigma^{\,2}}{iE_1\tau_\sigma}\langle\phi(r-R_i)|\exp\{i\boldsymbol{\tau}_\sigma(\mathbf{r}-\mathbf{R}_i)\}|\phi(\mathbf{r}-\mathbf{R}_i)\rangle \\ &= i(E_1\tau_\sigma/M^{1/2}\omega_\sigma^{\,2})\{1+(\tfrac{1}{2}a_0\tau_\sigma)^2\}^{-2}, \qquad \ldots\ldots\ (36) \end{aligned}$$

assuming that $\phi(\mathbf{r}-\mathbf{R}_i)$ describes a donor electron in an s level, with effective Bohr radius a_0. If the phonons are in phase at site i, the phases at site j will be random; thus $C_\sigma^{(j)}$ can be neglected. The sum over σ can be replaced by an intergration, using a Debye spectrum (cf. Lax and Burstein 1955).

$$S = \frac{3}{\omega_m^{\,3}}\int_0^{\omega_m}\frac{\omega}{\hbar}|C_\sigma^{(i)}|^2\omega^2\,d\omega = \frac{E_i^{\,2}\hbar^2}{Ma_0^{\,2}(k\Theta_D)^3}, \qquad \ldots\ (37)$$

where

$$\omega_m = k\Theta_D/\hbar$$

and Θ_D is the Debye temperature. Using constants appropriate to n-type germanium, $E_1 = 15$ eV, $a_0 = 44$ Å, $\Theta_D = 362°$K and $M = 2{\cdot}46\times10^{-22}$ g (twice the mass of the lattice atom since there are two atoms per unit cell), we find $S = 0{\cdot}10$. At finite temperatures,

$$S(T) = \frac{3}{\omega_m^{\,3}}\int\frac{\omega}{\hbar}|C_\sigma^{(i)}|^2\coth\frac{\hbar\omega_0}{2kT}\omega^2\,d\omega \simeq S(0)\left\{1+3\left(\frac{T}{\Delta}\right)^2\right\} \quad . \quad (38)$$

when T/Δ is small. Here, if v is the velocity of sound in the crystal, $k\Delta = \hbar v/a_0$ is the energy of a phonon with wavelength comparable to a_0; these phonons interact most effectively with the localized electrons. For germanium, using $v = 5{\cdot}3\times10^5$ cm sec^{-1}, we find $\Delta \simeq 9°$K. Hence $S(T)$ is small throughout the temperature region in which impurity conduction is important ($T \lesssim 5°$K) in germanium.

The transition probability W_{ij} can be calculated in a similar manner to that for the strong coupling case. We study the motion of a single electron in a random lattice of $N+1$ positive and N negative centres. This is equivalent to n-type impurity conduction, in which all except one electron are assumed to be stationary. The electron Hamiltonian is

$$\mathscr{H}_e = p^2/2m^* + \sum_\nu v(\mathbf{r}-\mathbf{R}_\nu) - \sum_i v(\mathbf{r}-\mathbf{R}_i),$$

where

$$v(\mathbf{r}-\mathbf{R}) = e^2/\kappa\,|\mathbf{r}-\mathbf{R}|$$

and $\mathbf{R}_\nu$ is the coordinate of a negatively charged centre, $\mathbf{R}_i$ that of a positive one. The overlap between electron wave functions centred on adjacent positive centres must be treated carefully, since the conductivity will vanish in the absence of overlap. Let

$$u_i = u(\mathbf{r}-\mathbf{R}_i)$$

be the wave function of an electron on an isolated donor, and

$$\langle u_i\,|\,u_j\rangle = \Delta_{ij}\ (\neq 0)$$

be the overlap integral. Then (Löwdin 1956), the funtions

$$\phi_i(r) = \sum_j \Delta_{ij}^{-1/2}\, u_j(r) \qquad \ldots\ldots\ (39)$$

form an orthogonal set. The sum is over all positive centres. Matrix elements of $\mathscr{H}_e$ in terms of the ϕ_i are

$$\langle \phi_i\,|\,\mathscr{H}_e\,|\,\phi_j\rangle = \epsilon^0\,\delta_{ij}\,V_{ij},$$

where ϵ^0 is the energy of an isolated donor electron, and

$$V_{ij} = \sum_{k,l} \Delta_{ik}^{-1/2}\, V_{kl}^{\,0}\, \Delta_{lj}^{-1/2}. \qquad \ldots\ldots\ (40)$$

$V_{kl}^{\,0}$ is the matrix element of the electrostatic potential of all centres except the lth positive one; so that

$$V_{kl}^{\,0} = \langle u_k(\mathbf{r})\,|\,\sum_\nu v(\mathbf{r}-\mathbf{R}_\nu) - \sum_{i\neq l} v(\mathbf{r}-\mathbf{R}_i)\,|\,u_l(\mathbf{r})\rangle. \qquad \ldots\ (41)$$

Thus the ϕ_i describe electrons with energy ϵ_i,

$$\epsilon_i = \epsilon^0 + V_{ii},$$

the energy ϵ_{ii} arising from the potential of the surrounding randomly placed charged centres. Since the overlap Δ_{ij} is small,

$$\phi_i(r) \simeq u_i(\mathbf{r}) - \tfrac{1}{2}\,\Delta_{ij}\,u_j(\mathbf{r}).$$

Thus ϕ_i is mainly localized about the ith positive centre, and the off-diagonal elements V_{ij} lead to transitions of the electron from i to j.

The electron-lattice interaction can now be treated in exactly the same way as for the strong coupling case. We find that associated with the state $\phi_i(\mathbf{r})$ is a set of lattice vibrations described by the oscillator functions $X_{n_\sigma}(Q_\sigma^{(i)})$; n_σ is the number of phonons in mode σ. The displaced coordinates $Q_\sigma^{(i)}$ are by (28)

$$Q_\sigma^{(i)} = Q_\sigma + C_\sigma^{(i)},$$

where $C_\sigma^{(i)}$ is given by (36) to the lowest order in the overlap. The expression for the transition probability W_{ij} is therefore of the same form as (33), with V_{ij} replacing M;

$$W_{ij} = \frac{1}{\hbar^2} |V_{ij}|^2 \exp\{-S(T)\} \int \exp\{G(T,t) + i(\epsilon_i - \epsilon_j)t/\hbar\}\, dt, \quad . \quad (42)$$

where $G(T,t)$ is given by (34) and

$$\epsilon_i - \epsilon_j = V_{ii} - V_{jj} \simeq \langle u_i | v(\mathbf{r} - \mathbf{R}_i) | u_i \rangle - \Delta_{ij} \langle u_j | v(\mathbf{r} - \mathbf{R}_i) u_i \rangle$$

to the lowest order in the overlap and neglecting all but the nearest-neighbour centre j. Since $S(T)$ and therefore $G(T,t)$ is small, the single phonon contributions to W_{ij} dominate, and thus in (42) $\exp G$ can be replaced by $1+G$.

Assuming a Debye spectrum of lattice vibrations and performing the sum over modes σ as previously indicated for $S(T)$,

$$\begin{aligned} W_{ij} &\simeq \frac{1}{\hbar^2} |V_{ij}|^2 \int_{-\infty}^{\infty} dt \exp\{i(\epsilon_i - \epsilon_j)t/\hbar\} \frac{3}{2\omega_m{}^3} \int_{-\infty}^{\infty} d\omega \exp(i\omega t) \\ &\quad \times \frac{\omega^3}{2\hbar} |C_\sigma^{(i)} - C_\sigma^{(j)}|^2 \left(1 + \coth \frac{\hbar\omega_\sigma}{2kT}\right) \\ &= 3\pi |V_{ij}|^2 \frac{E_1{}^2}{M\hbar v^2 (k\Theta_D)^3} \frac{\epsilon_i - \epsilon_j}{[1 + \{(\epsilon_i - \epsilon_j)a_0/2v\hbar^{s2}]^4} \\ &\quad \times \left\{1 + \coth \frac{\epsilon_i - \epsilon_j}{2kT}\right\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43) \end{aligned}$$

We have used the value of $C_\sigma^{(i)}$ calculated in (36).

The two phonon processes, which involve an integration over $\{G(T,t)\}^2$ give a contribution $\sim 10^{-2}$ times the single phonon transition rate (cf. eqn. 35).

§ 8. Calculations of Impurity Conduction

The theory of impurity conduction has been studied by a number of authors: Aigrain (1954), Baltensperger (1953), Conwell (1956), Erginsoy (1950, 1952), Mott (1956), and Kasuya and Koide (1958). In this section we shall outline the recent work of Miller and Abrahams (1960) and Twose (1959) on conduction in the low concentration region in germanium and silicon. The former authors use the following method.

(*a*) The probability per unit time W_{ij} that an electron jumps to a neighbouring vacant site is calculated, as a function of the separation between the two sites. A phonon is emitted or absorbed to conserve energy, as explained in the previous section.

(*b*) Equations for the rate of change of the probability that an electron occupies a given site and the net current flow, in an electric field, are written down and shown to be equivalent to Kirchoffs' Laws for charge flow in a three-dimensional random resistance network. Each link in the network corresponds to two impurity centres; the link impedance is inversely proportional to W_{ij}, the transition rate between the centres.

(*c*) The network resistivity is computed, assuming that it arises from non-intersect chains of impedances taken in parallel, each link in the chain being chosen in a suitable manner.

In evaluating W_{ij} in the previous section we assumed a simple hydrogenic form for the donor electron wave function,

$$u_i(r) = (\pi a_0^3)^{-1/2} \exp(-r/a_0).$$

The correct effective mass wave functions have been described in § 3. Using these wave functions, and averaging over the possible directions in the crystal of the neighbouring vacant site, these authors found

$$\langle W_{ij} \rangle_{\text{ave}} = \frac{E_1^2}{2\pi\rho_0 v^5 \hbar^4} U^2 \Delta \{\coth(\Delta/2kT) + 1\},$$

where

$$U^2 = (2e^2/3\kappa a^2)^2 (\pi a/4\alpha R)^{1/2} R^2/n \exp(-2R/a).$$

Here ρ_0 is the density and v the velocity of sound in the crystal, and E_1 the deformation potential constant. The coth describes the number of phonons present with energy $|\Delta|$, where Δ is the difference between the energies of the two centres. R is the distance between the two centres and $\alpha = a^2/b^2 - 1$; a, b are the Bohr radii and n the number of conduction band minima (§ 3).

If f_i is the probability that centre i is occupied by an electron,

$$\partial f_i/\partial t = \sum_j \{W_{ij} f_i (1 - f_j) - W_{ji} f_j (1 - f_i)\}. \quad . \quad . \quad . \quad . \quad (44)$$

This determines the equilibrium distribution. As in § 2,

$$f_i = 1/[1 + \exp\{(\epsilon_i - \zeta)/kT\}],$$

where ζ is the Fermi energy. From considerations of detailed balancing, we have

$$W_{ij} \exp(\epsilon_j/kT) = W_{ji} \exp(\epsilon_i/kT).$$

In the presence of a small electric field F in the x direction, a different equilibrium distribution f_i' of electrons is formed, determined by a steady-state condition of the form (44) but with f_i replaced by f_i', and the transition probabilities which are altered by the field now obey

$$W_{ij} \exp\{(\epsilon_j - e x_j F)/kT\} = W_{ji} \exp\{(\epsilon_i - e x_i F)/kT\}; \quad . \quad . \quad (45)$$

x_i is here the x coordinate of the ith centre. f_i' can be written formally (assuming local equilibrium), in terms of a parameter E_i, as

$$f_i' = 1/[1 + \exp\{(E_i - \zeta)/kT\}].$$

Substituting this form for f_i' in the steady state condition, using the reciprocal relation (45) and taking terms linear in F only, we find

$$(1/kT) \sum_j \{(\epsilon_i - e x_i F) - (\epsilon_j - e x_j F)\} W_{ij} f_i (1 - f_i) = 0.$$

This is of the form of Kirchhoffs' first law for a network, with the term $\{\ \}$ corresponding to a potential drop across the 'link' ij, and the link impedance,

$$Z_{ij} \propto \{W_{ij} f_i (1 - f_j)\}^{-1}.$$

The problem is thus reduced to determining the impedance of an equivalent resistance network.

The solution of the network problem, in terms of chains of conducting elements, the chains being taken in parallel, is described in detail by Miller and Abrahams (1960). The final result for the resistivity can be written

$$\rho(T) = C(T)(r_D/a)\{1 + 18{\cdot}2(a/r_D)^{3/2}\} \exp\{1{\cdot}09(r_D/a)^{3/2} + (\epsilon_3 - \epsilon_e)/kT\} \qquad (46)$$

where

$$C(T) = 4{\cdot}55 \times 10^2 \, l_e(T)(\alpha/8)^{1/2} \kappa^2 n \rho_0 v^5 \hbar^4 a^3 / e^6 E_1^{\,2}$$

and

$$r_D = (3/4\pi N_D)^{1/3}; \quad \alpha = -1 + a^2/b^2.$$

Here a and b are the radii of the effective mass wave functions, n is the number of conduction band minima, ρ_0 the density and v the velocity of sound in the crystal. In the discussion above, the donor electron was assumed to occupy the ground state. l_e and ϵ_e contain the effects of excited states, which are shown to be unimportant ($l_e = 1$, $\epsilon_e = 1$) except in the case of antimony doped germanium (cf. § 3). Thus $\rho(T)$ is of the form

$$\rho(T) = \rho_0 (N_{maj}) \exp(\epsilon_3/kT).$$

In this calculation the whole of the compensation dependence is included in the activation energy ϵ_3, which is given by

$$\epsilon_3 = \zeta - 1{\cdot}35 \, \epsilon_A, \quad \epsilon_A = (e^2/\kappa)(4\pi N_A/3)^{1/3}.$$

The determination of the Fermi energy ζ (on the assumption that the energy difference $\epsilon_i - \epsilon_j$ between sites is due to nearest neighbour charged minority sites only) has been described in § 2. If the compensation is small ($K \lesssim 0{\cdot}2$)

$$\epsilon_3 = \epsilon_D - 1{\cdot}35 \, \epsilon_A = (e^2/\kappa)(4\pi N_D/3)^{1/3}(1 - 1{\cdot}35K^{1/3}).$$

For higher values of K, values of ϵ_3 are plotted in fig. 8. Although, as we have seen in § 4, the magnitude of ϵ_3 agrees well with experimental values, the temperature dependence of $\rho(T)$ is incorrect since it predicts that all curves plotting ρ against $1/T$ for the same concentration of majority impurity centres extrapolate to the same point at $T = \infty$.

Calculated values of ρ from (46) for antimony-doped germanium are compared with measured values in table 3. The agreement is satisfactory, considering that errors of at least 30% are possible from uncertainties in the effective Bohr radii and the deformation potential constant E_1.

Table 3. Resistivity ρ of germanium doped with antimony (Ω cm $\times 10^6$)

$N_D \times 10^{-15}$ cm^{-3}	1·6	2·3	3·0	5·2
ρ (2·5 °K)	$5{\cdot}8 \times 10^3$	$3{\cdot}2 \times 10^2$	50	5·6
ρ (calculated)	$1{\cdot}1 \times 10^4$	$5{\cdot}2 \times 10^2$	89	7·8

Twose (1959) has also calculated the impurity resistivity in the low concentration region by a somewhat different method. A density matrix approach was used, similar to that of Luttinger and Kohn (1958), using as a representation the product $\phi_i(r)X_n(Q^{(i)})$ of electron and oscillator functions, described earlier (§ 7). This leads to an average electron drift velocity $\langle v\rangle$ which is a sum of 'two centre' contributions v_{ij}. Each v_{ij} depends on the charge amplitudes of the electron on sites i and j, and hence $\langle v\rangle$ should be found from a self-consistent treatment of charge diffusion onto and away from a given centre. Instead of this self-consistent approach, the electron was assumed to occupy sites with a Boltzmann probability, leading to a two centre conductivity for N_e electrons of the form

$$\sigma_{ij} = N_e\, F(R_{ij}) f(T, \epsilon_{ij}).$$

This must then be averaged in suitable way over the random separations R_{ij} of the impurity centres. Here

$$F(R_{ij}) = \{3\pi e^2 E_1^{\,2}\hbar^2/Ma_0^{\,2}(k\Theta_D)^3\}|\langle\phi_i|v|\phi_j\rangle|^2,$$

where M is the mass of the atom, Θ_D the Debye temperature, and $\langle\ \rangle$ is the matrix element of the velocity between the localized electron wave functions on centres i and j. The temperature and energy dependence of σ is contained in the equation

$$N_e f(T, \epsilon_{ij}) = \frac{(\epsilon_i-\epsilon_j)/2k\Delta}{\{1+[(\epsilon_i-\epsilon_j)/2k\Delta]^2\}^4}\frac{1}{kT}\left\{1+\coth\frac{\epsilon_i-\epsilon_j}{2kT}\right\}N_e \exp(-\epsilon_i/kT)/\sum\exp(-\epsilon_j/kT). \quad . \quad . \quad . \quad . \quad . \quad (47)$$

ϵ_i is here the energy of an electron on site i, $k\Delta = hv/a_0$ is the energy of a phonon of wave length comparable to the Bohr radius a_0 of the localized electron, where v is the velocity of sound. Since $\epsilon_i - \epsilon_j$ is independent of R_{ij} except for very small separations, the dependence of σ_{ij} on separation is almost entirely contained in the term $|\langle\phi_i|v|\phi_j\rangle|^2$.

In the case of small compensation, $f(T, E)$ can be approximated by use of the trapping model (§ 2). The electron jumps between the 'free' sites, where $\epsilon_i \simeq \epsilon_j$, and the Boltzmann term in (47)

$$N_e \exp(-\epsilon_i/kT)/\sum\exp(-\epsilon_j/kT),$$

leads to the number n_f of free carriers,

$$n_f = (N_{\min}N_{\mathrm{maj}})^{1/2}\exp(-\epsilon_t/2kT),$$

where ϵ_t is the energy difference between free and trap sites. Then

$$f(T, \epsilon) = n_f/k\Delta.$$

The conductivity becomes

$$\sigma = n_f e\,\mu_f,$$

where

$$\mu_f = \{3\pi e E_1^{\,2}\hbar/Ma_0 v(k\Theta_D)^3\}|\langle\phi_i|v|\phi_j\rangle|^2.$$

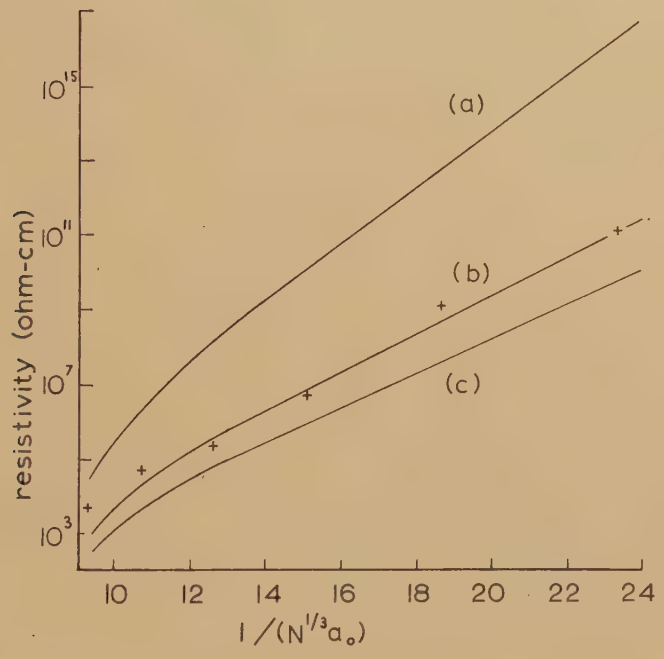
Thus the mobility is independent of temperature.

Assuming hydrogen-like wave functions for the localized electrons in evaluating the velocity matrix element, it was found that

$$\sigma_f = \frac{\pi e^6 E_1^2}{3\hbar\kappa^2 M v^2 (k\Theta_D)^2} \nu^2 (\nu M)^2 \exp(-2\nu)(N_{maj}N_{min})^{1/2} \exp(-\epsilon_t/2kT) \qquad (48)$$

where $\nu = R/a_0$. When averaging σ_f over the random centre distribution, care must be taken not to over emphasize the contribution from pairs of very close centres, since the electron will tend to be trapped on such a pair rather than move on through the lattice. One rough estimate of the average conductivity $\bar{\sigma}$ is obtained by assuming a constant impurity separation. Figure 17 shows that the observed resistivity is best fitted by $1/\bar{\sigma}$ if

Fig. 17



The resistivity for n-type germanium calculated from (48) assuming a constant average centre separation ν of (*a*) $N^{-1/3}$, (*b*) $0{\cdot}7\ N^{-1/3}$ and (*c*) $0{\cdot}62\ N^{-1/3}$ The crosses are experimental points (Fritzsche 1958).

$\nu \sim 0{\cdot}7/N^{1/3}a_0$, implying that conductivity takes place preferentially along chains of impurities whose spacing is less than average. In evaluating $\bar{\sigma}$ in fig. 17 we have set $\frac{1}{2}\epsilon_t$ for ϵ_3, the observed activation energy.

Another averaging procedure, suggested by Pippard, is the following. Suppose that a pair of centres, separated by a distance R and with ' two centre ' conductivity $\sigma(R)$, can be replaced by a sphere of radius R and a uniform conductivity $\sigma(R)$. Let this sphere be imbedded in a medium

with the average conductivity $\bar{\sigma}$ of the disordered lattice as a whole. When a field F is applied, the field inside the sphere becomes

$$F(R) = 3\bar{\sigma}F/\{2\bar{\sigma} + \sigma(R)\}.$$

The current density inside the sphere is $F(R)\,\sigma(R)$. If the probability of finding a sphere with radius between R and $R+dR$ is $P(R)\,dR$, then the average current density $\bar{j}$ is given by an integral equation

$$\bar{j} = \bar{\sigma}F = \int_0^\infty P(R)\,\frac{3\bar{\sigma}\sigma(R)}{2\bar{\sigma}+\sigma(R)}\,F\,dR.$$

For a random impurity distribution the probability function is

$$P(R) = 4\pi N_{\mathrm{maj}}\,R^2 \exp\left(-4\pi N_{\mathrm{maj}}\,R^3/3\right).$$

The equation was solved numerically.

Calculated values of the resistivity ρ are compared with measured values ρ_{exp} for antimony doped germanium (Fritzsche 1958), at a temperature of 2·5°K, in table 4. The observed activation energy ϵ_3 has been used to estimate ϵ_1, through the relation $\epsilon_3 = \frac{1}{2}\epsilon_1$ given by the trapping model (§ 2).

Table 4

N_{D} cm^{-3}	K	ρ (Ω . cm)	ρ_{exp}(Ω . cm)
$9{\cdot}3\times10^{14}$	0·012	$6{\cdot}5\times10^{11}$	$7{\cdot}1\times10^{11}$
$1{\cdot}6\times10^{15}$	0·014	$5{\cdot}2\times10^{9}$	$6{\cdot}3\times10^{9}$
$2{\cdot}3\times10^{15}$	0·010	$7{\cdot}4\times10^{8}$	$3{\cdot}2\times10^{8}$
$3{\cdot}0\times10^{15}$	0·010	$1{\cdot}8\times10^{8}$	$7{\cdot}1\times10^{7}$
$5{\cdot}2\times10^{15}$	0·010	$2{\cdot}3\times10^{7}$	$5{\cdot}6\times10^{6}$
$8{\cdot}5\times10^{15}$	0·014	$1{\cdot}4\times10^{6}$	$1{\cdot}4\times10^{6}$
$1{\cdot}3\times10^{16}$	0·08	$6{\cdot}6\times10^{4}$	$2{\cdot}2\times10^{5}$

Both the above calculations are based essentially on choosing an average transition rate W_{ij} of the electron between two centres. The averaging procedure is critical, since W_{ij} increases exponentially with decreasing centre separation. As mentioned earlier, the large transition rate between two centres spaced closer than the average does not imply there is a correspondingly large contribution to the average conductivity, since the electron may have difficulty in proceeding from j to more distant centres. The charge on j will then build up by an amount which depends on the charge distribution on surrounding centres and on the applied electric field, until the forward transition rate $i \rightarrow j$ is balanced by the back transition rate $j \rightarrow i$.

Another point is that if f_i is the probability that centre i is occupied, the probability that the nearest neighbour j is unoccupied will be larger than $(1-f_i)$, due to Coulomb repulsion between the electrons. Miller (private communication) points out that this neglect of correlations between electrons may be the cause of the incorrect compensation dependence of his results. To conclude, a rigorous calculation of the conductivity will

involve setting up a transport equation in which the charge distribution in the presence of an electric field is treated self-consistently, and with carrier–carrier correlation taken into account. A part of this programme has been completed by one of us (W.D.T.) in that the correct transport equation has been derived in the approximation that correlations are neglected; as might be expected the equation takes the form of a generalized Einstein relationship. The inclusion of the important self-consistency and correlation effects is being investigated.

Part II. The Transition to a Metallic Form of Conductivity

§ 9. Introduction

It is a property of silicon, germanium and of many other extrinsic semiconductors that, as the concentration N of impurities increases, the activation energy ϵ_3 for conduction in the temperature range for impurity conduction decreases and, at a critical value N_c, vanishes. For values of N greater than N_c the resistivity and Hall constant are roughly independent of temperature down to the lowest temperatures for which measurements have been made.

In a number of papers one of us (Mott 1949, 1952, 1956, 1957, 1961) has given arguments to suggest that a sharp transition from a metallic to a non-metallic state must occur for a *crystalline* array of atoms as the distance between them is increased. In the paper published in 1956 it was suggested that the transition described in this paper for a random array of centres is of this type, the sharpness being lost because of the disordered arrangement. We consider that this is the correct explanation. However, another explanation is certainly possible, namely that the transition is one due essentially to the disordered lattice and occurs at the concentration at which the states for a single carrier become localized. We shall have to examine the evidence that this is not so at any rate for low K.

We shall not repeat the arguments summarized by Mott (1961) that this transition occurs for a crystalline array, remarking only that, while an exact calculation of the transition concentration has not proved possible, it should occur at a constant value of about 3 of the constant

$$\lambda = (3/4\pi N_{\mathrm{maj}})^{1/3}/a_0 = r_s/a_0,$$

where a_0 is the Bohr radius of the centre.

We shall now summarise what we think happens in the disordered lattice in the two limiting cases, $K \ll 1$ and $1-K \ll 1$, where K is degree of compensation.

(a) Low compensation, $K \ll 1$. In the region of low concentration, $(N/N_c \ll 1)$, the carriers are holes, bound to minority centres by a Coulomb field $e^2/\kappa r^2$. As N increases, therefore, the holes cannot become free, because a Coulomb field always leads to bound states. The transition concentration N_c should increase slightly with compensation K, because

some of the majority centres are empty and so the amount of overlap between localized centres is decreased (see below).

(*b*) High compensation, $1-K \ll 1$. Here we have at present little experimental evidence. The theoretical predictions are the following. As the concentration N of majority carriers is increased, a value will be reached for which the electrons are no longer in bound states (§ 2). Thus n electrons per unit volume (where $n=(1-K)N$) move in the random field due to fixed positive and negative charges; this random field is thus not strong enough to give bound states. There is thus the possibility of a 'crystallization' of electrons as predicted by Wigner (1938). This is discussed by Mott (1961). We are convinced that, whether or not the electrons behave like a classical liquid or a condensed electron gas, their conductivity will be high and that in this case any transition which may be observed will be due to a transition from bound to free states for a *single* electron.

The remainder of this paper will deal with the case of small or moderate value of K, for which most of the experimental work has been done.

§ 10. Dependence of the Transition Concentration or Degree of Compensation

If the transition were due essentially to the properties of the random lattice, we should expect it to depend very sensitively on K, since K determines the random field. If on the other hand the transition is due simply to overlap between occupied centres, we should expect a variation of λ according to the formula

$$\lambda = \text{const}\,(1-K)^{1/3}.$$

Some experimental support for this compensation-dependence of λ is given by measurements of the acceptor separation r_s at which the activation energy ϵ_3 disappears in p-type gallium-doped germanium (Fritzsche and Cuevas 1960 b). The results are summarized in table 5, and are shown in fig. 18.

Table 5

r_s (Å)	127	110	98
K	0·04	0·4	$0{\cdot}4 < K < 0{\cdot}7$

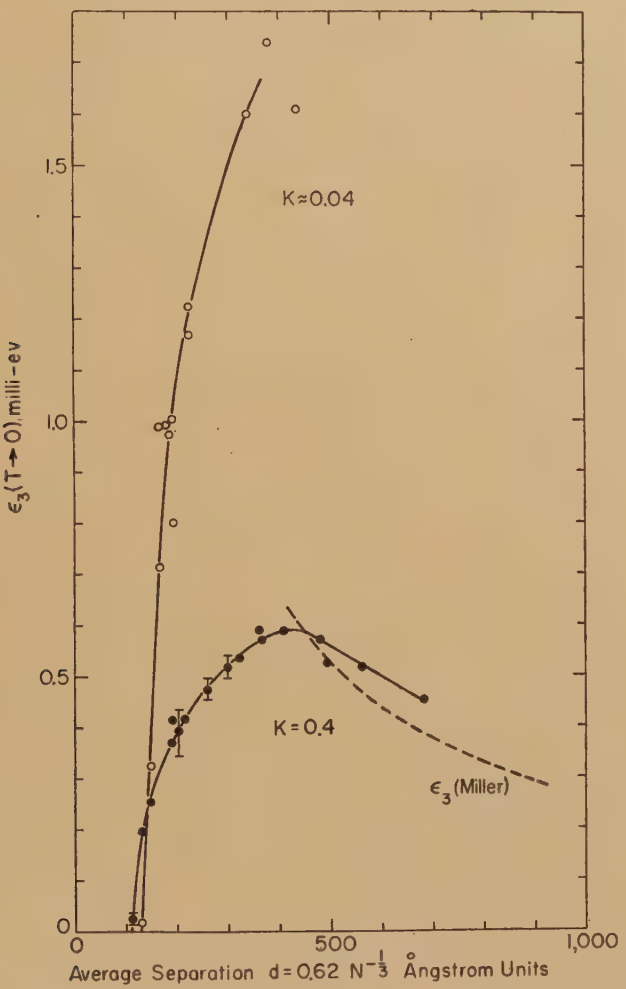
The critical separation decreases as K increases. We also note that the ratio of the separations for K equal to 0·04 and 0·4 is 1·18, while that calculated from (48) is 1·12.

The experimental value of λ is not clear, since the gallium wave functions are not accurately known. A variational calculation shows that at large distances from the impurity the wave function is a sum of two exponentials. One has a Bohr radius $a_1 = 35 \pm 7$ Å, the other $a_2 = 89 \pm 2$ Å (Miller and Abrahams 1960). Then for $K = 0{\cdot}04$,

$$\lambda_1 = r_s/a_1 = 3{\cdot}6, \quad \lambda_2 = r_s/a_2 = 1{\cdot}4.$$

Since the relative amplitudes of the two exponentials are not known, we can only deduce that $1{\cdot}4 < \lambda < 3{\cdot}6$. However, λ is probably closer to 3·6 than 1·4. One reason for this is that in the alkali metals (where the conductivity is of course 'metallic') the inter-electron separation in units of the Bohr radius range from 3·22 (lithium) to 5·57 (caesium).

Fig. 18



Activation energy ϵ_3 of impurity conduction of transmutation-doped p-type germanium as a function of average impurity separation for $K = 0{\cdot}4$ and $K = 0{\cdot}04$ (Fritzsche and Cuevas 1960 b).

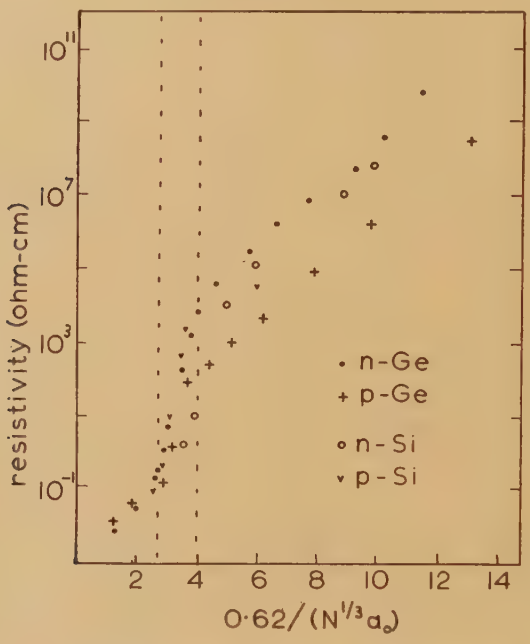
§ 11. Concentration at which the Transition Occurs

In fig. 19 we show the resistivity at 2·5 °K of weakly compensated samples of n- and p-type silicon and germanium. This is plotted against λ, in other words $(3/4\pi N)^{1/3}/a_0$ or r_s/a_0, where r_s is the mean distance between

impurities. The Bohr radius a_0 used here is obtained by assuming the hydrogen atom model for an impurity centre, and adjusting a_0 to fit the observed ionization energy of the impurity centre. The activation energy ϵ_3 drops rapidly in the transition region (shown in fig. 19 between the vertical dotted lines) and occurs for all substances roughly when $\lambda \simeq 3$, or $N^{1/3}a_0 \simeq 0{\cdot}2$.

Roughly the same value is deduced by Dewald (1960) from the measurements of Thomas (1959) for zinc oxide.

Fig. 19



Variation of resistivity with average impurity separation for weakly compensated specimens of germanium and silicon. The transition region is enclosed by the dotted lines. The activation energy ϵ_3 for impurity conduction vanishes at the small separation end of the transition region.

McIrvine (1960), by examining a number of semiconductors, deduces an empirical formula which relates N_c to the static dielectric constant ϵ. However, for semiconductors that are not elements, it is not clear what value of the dielectric constant one should take, the high frequency value, the static value or some mean between them; this point needs further investigation before an assessment of the meaning of McIrvine's formula can be given.

What is happening in the transition region is not clear. We have seen (fig. 2) that the ratio of the Hall constant to $1/N_{\mathrm{maj}}$ drops in this region. This suggests that the material is inhomogeneous, due to fluctuations in the density of centres, and that there are small metallic regions in series with

non-metallic ones, whose relative volumes depend on temperature. If the Hall effect in the non-metallic regions is small or non-existent, one would expect, as the temperature is lowered and the volume of the non-metallic regions increases, a drop in the observed Hall constant. This is a possible explanation of the effect.

It should however be pointed out that Read and Katz (1960) have observed a Hall effect for ionic conduction in KCl; the hopping process of a point defect from one centre to another is not dissimilar to that described here.

§ 12. Resistivity in the Region of Metallic Conduction

The conductivity of a metal may be written

$$\sigma = Ne^2 l/m^* v, \quad \ldots \ldots \quad (49)$$

where l, v are the mean free path and velocity at the surface of the Fermi distribution. Moreover, for a degenerate electron gas,

$$m^* v/\hbar = 2\pi(3N/8\pi)^{1/3}.$$

Also it is convenient to write

$$l = p/N^{1/3},$$

so that p is the number of interatomic distances on a mean free path. Then (49) becomes

$$\sigma = N^{1/3} e^2 p/\hbar(3/\pi)^{1/3},$$

or in $\text{ohm}^{-1}\,\text{cm}^{-1}$, if N is in particles per cubic angstrom,

$$\sigma = 7 \times 10^3 p N^{1/3},$$

For the specimens with highest conductivity shown in fig. 4, N is 10^{-6} and the observed conductivity is $1{\cdot}5 \times 10^2$, so p is about 2. This is shown also in the following table (Fritzsche 1960, private communication) for n-type germanium, in which v is deduced from (50) and hence the mean free path from the observed resistivity p. It is of course assumed that each centre contributes a free electron, so $N = N_D$. The last column shows the mean distance between centres. The table also shows that l tends to remain constant as N decreases, so p apparently drops below unity.

Table 6.

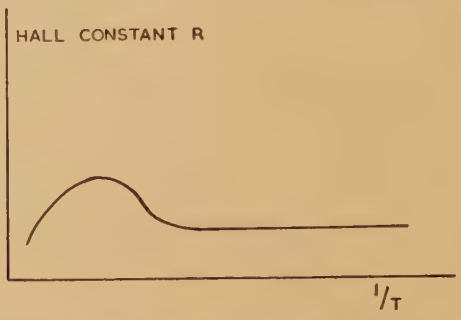
ρ (obs, Ω . cm)	N_D (cm^{-3})	v ($\text{cm/sec} \times 10^{-6}$)	l ($\text{cm} \times 10^8$)	$0{\cdot}62\ N^{-1/3}$ ($\text{cm} \times 10^8$)
0·02	2×10^{17}	6	62	106
0·0095	6×10^{17}	8·6	63	73
0·0067	10^{18}	10	64	62
0·0033	3×10^{18}	15	63	43
0·00157	10^{19}	22	59	29
0·001	2×10^{19}	28	58	23
0·00066	4×10^{19}	35	55	18

We should expect l to be given by a formula of the form

$$1/l = N \int I(\theta)\,(1-\cos\theta)\,2\pi \sin\theta\, d\theta, \quad . \quad . \quad . \quad . \quad (51)$$

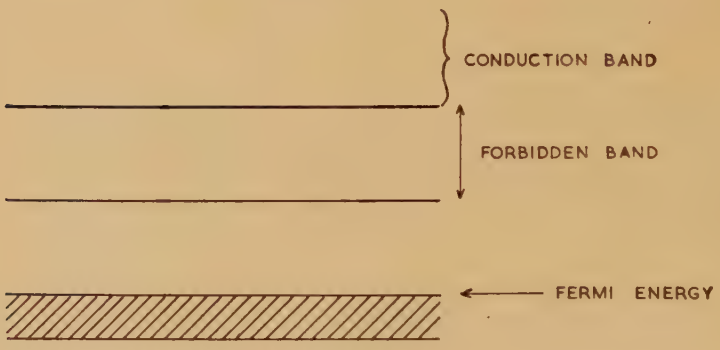
where $I(\theta)$ is the differential cross section for scattering by each centre.

Fig. 20



Hall constant near transition concentration in metallic region.

Fig. 21



An impurity band for concentration near the transition point.

That, in the metallic region, the integral is of the order of $N^{-2/3}$ is perhaps not unexpected for a totally disordered lattice. The constancy of l, however, means that the integral is proportional to $1/N$, and this is unexpected. A crude application of Born's approximation would suggest that the scattering potential energy function of each centre was of order $e^2/\kappa R\,(4\pi R^3/3 = N)$ extending over a volume R^3, so that

$$I(\theta) = \left| \frac{1}{4\pi} \frac{2m}{\hbar^2} \int \exp\{i\mathbf{k} - \mathbf{k}' \,.\, \mathbf{r}\} \frac{e^2}{\kappa R d\tau} \right|^2,$$

the integral being over a sphere of radius R and $|\mathbf{k}-\mathbf{k}'|=2k\sin\frac{1}{2}\theta$. Since kR is independent of N. this must be equal to a numerical factor multiplied by $R^4/a_0{}^2$, where a_0 is the hydrogen radius $\hbar^2\kappa/m^*e^2$. Thus l should vary as R^{-1}, i.e. as $N^{1/3}$.

The observed constancy may perhaps be due to the non-Coulomb part of the field near the impurity atom, the relative importance of which will increase with N.

A point about the conductivity near the transition point may be pointed out. The Hall constant plotted against $1/T$ shows the behaviour sketched in fig. 20 (cf. also fig. 4). This suggests that there is an 'impurity band', formed from the impurity wave-functions, and not yet overlapping the conduction band. If the impurity wave functions are hydrogen-like, it is easy to estimate its width. This is, using the approximation of tight binding (Mott and Jones 1936, Chap. III, Mott 1957)

$$B=6W_0(1+\lambda')\exp(-\lambda')$$

where W_D is the ionization energy of a centre. The assumption is made that the centres are arranged in a simple cubic lattice of side b, and λ' is thus b/a_0. Values of the ratio of B to W_0 are

λ'	B/W_0
2	3·0
3	1·2
4	0·54
5	0·24

Only values less than unity have any meaning in this approximation. The experimental value of λ at the transition point is *c.* 3, and if we set

$$\lambda'=(4\pi/3)^{1/3}\lambda=1/Na_0,$$

a value of about 5 is appropriate for λ'. Near the transition concentration, then this rough estimate suggests, as do the experiments, that the impurity band does not yet overlap the conduction band.

ACKNOWLEDGMENTS

We are grateful to several of our colleagues with whom we have discussed these problems, particularly Dr. M. H. Cohen and Dr. H. Fritzsche.

APPENDIX

THE EXISTENCE OF BOUND STATES IN THREE DIMENSIONS

We give here an outline of an extension due to one of us (W.D.T.) of the work of Anderson (1958) showing that the single electron states are bound in an impurity lattice when the average impurity separation greater than a critical value ; the degree of compensation is assumed close to unity.

We study the wave function of an electron in a lattice of positive and negative fixed point charges which are distributed at random in a dielectric medium. There is a density N of positive charges, and the total number of positive charges exceeds the number of negative charges by one. Let $\phi_i(r)$ be the wave function of the electron localised on the ith positive site. For simplicity, we here assume that ϕ_i is a hydrogen-like s function of the form

$$\phi_i(r) = (\pi a_0^3)^{-1/2} \exp(-r/a_0), \qquad \text{(A 1)}$$

and that ϕ_i, ϕ_j are orthogonal; we thus neglect the small overlap (ϕ_i, ϕ_j). (A more exact treatment can be given in terms of the orthogonalized ϕ_i of § 7, (eqn. 39). The wave function of the electron at time t can be written

$$\psi(t) = \sum a_i(t)\, \phi_i(r).$$

We assume the electron is initially localized on atom i, so that $a_i(0) = 1$ and $a_j(o) = 0$, and study the variation of $a_i(t)$ with t using the Schrodinger equation

$$i\hbar\, \partial a_i/\partial t = E_i a_i + \sum V_{ij} a_j. \qquad \text{(A 2)}$$

Here E_i is the energy of the electron on site i in the coulomb potential of the surrounding positive and negative charges. V_{ij} is the matrix element of the potential of all charged centres except the jth positive one:

$$V_{ij} = \left\langle \phi_i \left| \sum_\nu \frac{e^2}{\kappa |\mathbf{r} - \mathbf{R}_\nu|} - \sum_{k=j} \frac{e^2}{\kappa |\mathbf{r} - \mathbf{R}_k|} \right| \phi_i \right\rangle,$$

$\mathbf{R}_k$, $\mathbf{R}_\nu$ are the positions of the positive and negative centres respectively.

It is convenient to study the Laplace transform of (A 2), defining

$$f_i(s) = s \int_0^\infty \exp(-st)\, a_i(t)\, dt.$$

Thus s is an inverse time, and

$$\lim_{t\to\infty} a_i(t) = \lim_{s\to 0} f_i(s).$$

Then, it follows from the work of Anderson (1958) that the behaviour of $a_i(t)$ at large times depends on the convergence of an infinite perturbation series

$$V_c(s) = \sum_k V_{ik}(1/d_k)\{V_{ki} + \sum_m V_{km}(1/d_m) V_{mi}\}, \qquad \text{(A 3)}$$

where

$$d_k = i\hbar s - E_k. \qquad \text{(A 4)}$$

For, if t is large, Anderson shows that

$$a_i(t) \sim \exp(-t/\tau_i) \exp\{-i(E_i - \Delta_i)t/\hbar\},$$

where

$$\lim_{s\to 0+} V_c(s) = -\Delta_i + i\tau_i\hbar.$$

The imaginary part of V_c corresponds to a decay time, and the real part to an energy shift. Hence, if the Im $(V_c(s))$ vanishes as Re $(s) \to 0^+$, the amplitude of the electron on site i remains finite as $t \to \infty$, i.e. the electron is bound

to that site and has energy $E_i - \Delta_i$. If the imaginary part does not vanish, then after a sufficiently long time the electron will have diffused completely away from its initial position; bound states are in general not possible at any centre i.

$V_c(s)$ should be evaluated when

$$\mathrm{Im}\,(s) = (E_i - \Delta_i)/\hbar.$$

Thus in (A 6)

$$d_j = i\hbar s + (E_i - \Delta_i - E_j),$$

where s is real. In the following we neglect the small energy correction Δ_i, and write

$$E_{ij} = E_i - E_j.$$

The quantities V_{ij} and E_{ij} entering V_c can be written

$$V_{ij} = \left\langle \phi_i \left| v_{ij}(r) - \frac{e^2}{\kappa|\mathbf{r} - \mathbf{R}_i|} \right| \phi_j \right\rangle \quad . \quad . \quad . \quad . \quad (\mathrm{A}\,5)$$

$$E_{ij} = v_{ij}(R_i) - v_{ij}(R_j),$$

where

$$v_{ij}(r) = \sum_{\nu} e^2/\kappa\,|\mathbf{r} - \mathbf{R}_\nu| - \sum_{k \neq i,j} e^2/\kappa\,|\mathbf{r} - \mathbf{R}_k|$$

is the Coulomb potential at r of all the randomly placed positive and negative centres except the ith and jth positive ones.

The dominant contributions to V_c come from small values of E_{ij} and large values of V_{ij}. From (A 1),

$$V_{ij} \sim \exp(-R_{ij}/a_0),$$

where R_{ij} is the separation of i and j; thus V_{ij} is largest when i and j are near-neighbour centres. In general, for two near neighbours, E_{ij} is small when there are no close centres to i and j. Then, it is a fair approximation to neglect $v_{ij}(r)$ in (A 5), taking

$$V_{ij} \simeq \left\langle \phi_i \left| -\frac{e^2}{\kappa|\mathbf{r} - \mathbf{R}_i|} \right| \phi_j \right\rangle = -\frac{e^2}{\kappa a_0}\left(\frac{R_{ij}}{a_0} + 1\right)\exp(R_{ij}/a_0).$$

Also we note that in this approximation V_{ij} and E_{ij} can be treated independently, since V_{ij} depends on centres i and j only, while E_{ij} depends on the potential of all other centres.

Let

$$V_c = -\Delta(s) + i\hbar s\,X(s), \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{A}\,6)$$

where Δ, X and s are real. Then, in the limit $s \to 0^+$,

$$X = \sum_k (V_{ik}/E_{ik})\{V_{ki}/E_{ki} + \sum_m V_{km}V_{mi}/E_{km}E_{mi} + \ldots.\}. \quad . \quad (\mathrm{A}\,7)$$

We treat the V_{ij} and E_{ij} as independent random variables, obtain a probability distribution for X and show that the probability that X diverges to infinity is zero if the density N of charged centres is less than a critical value N_c. Then (cf. A 6),

$$\lim_{s \to 0} \mathrm{Im}\, V_c(s) = 0,$$

and the electron remains localized about site i.

The steps are the following:

(i) Anderson shows that the first term in (A 7),

$$X^{(1)} = \sum_k |V_{ik}|^2 / E_{ik}^2,$$

is convergent in the above sense if $V_{ij}(R_{ij})$ falls off faster than $1/R_{ij}^{3+\epsilon}$ ($\epsilon < 0$). This condition is obviously satisfied for our case.

(ii) We must now show that the complete series V is convergent when $N < N_c$. Let

$$X = \sum_L (S_L), \quad S_L = \sum (\pm T_L)$$

where S_L is the sum of all terms of length L; the 'length' is the number of times V/E appears in the term. The sign of both S_L and T_L is random, because of the random sign of the energy denominators E_{ij}. The number of terms of length L, with value T_L in the range $T_L - T_L + dT_L$, is found to be of the form

$$n(T_L)\, dT_L = [f(N)]^L L(T_L)\, dT_L / T_L^2,$$

where $f(N)$ increases with N. If $L(T)$ increases, or decreases no more rapidly than $T^{-1/2}$, then the probability distribution of S_L is of the form

$$P(S_L) dS_L \sim [f(N)]^L L(S_L)\, dS_L / S_L,$$

for values of the sum S_L greater than or of order the most probable value. Since we are interested only in the convergence of X, we consider the case in which L is large. Then N_c is determined by

$$[f(N)]^L L(1) = 1. \qquad \text{(A 8)}$$

For, if $N < N_c$, $f(N) < f(N_c)$ and the probability of obtaining $S_L = 1$ is smaller by a factor of order e^{-L}. The number of S_L in X increases as L, and the probability is $1 - e^{-L}$ that each term is less than unity. Hence X must almost always converge when $N < N_c$.

The detailed derivation of $n(T)$ will not be given here. It can be obtained by the methods in Anderson's paper, with the following assumptions.

(*a*) The probability distribution of the energy differences E_{ij} is of the form

$$\begin{aligned} P(E_{ij}) &= W^{-1} \quad \text{if} \quad -\tfrac{1}{2}W \leqslant E_{ij} \leqslant \tfrac{1}{2}W \\ &= 0 \qquad \text{if} \quad |E_{ij}| > \tfrac{1}{2}W, \end{aligned} \qquad \text{(A 9)}$$

(*b*) $$V_{ij} = V_0 \exp(-R_{ij}/a_0), \quad V_0 = e^2/\kappa a_0. \qquad \text{(A 10)}$$

The number of terms V_{ij} with values in the range $V - V + dV$ is obtained by assuming that the number of positive sites j at distance R_{ij} from i is

$$n(R_{ij}) dR_{ij} = 4\pi N R_{ij}^2\, dR_{ij}.$$

Then we obtain

$$N(T) dT = \left[\frac{16^3 \pi N a_0^3 V_0}{3^3 W}\right]^L \{1 + \ln(T/4L) - \tfrac{1}{4}\ln(2V_0/W)\}^L dT/T^2; \qquad \text{(A 11)}$$

W can be determined as follows. We write

$$E_{ij} = \sum_k Z_k$$

where

$$Z_k = \pm \frac{e^3}{k}\left(\frac{1}{R_{ik}} - \frac{1}{R_{jk}}\right),$$

and k is any one of the positive or negative centres. Treating Z_k as a step in a random walk problem, the probability distribution of E_{ij} is (Chandrasekhar 1943)

$$P(\epsilon) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(iE\tau)\exp\left[-4N\int \sin^2\left\{\frac{e^2\tau}{2\kappa}\left(\frac{1}{R_{jk}} - \frac{1}{R_{jk}}\right)\right\} d^3R_{ik}\right] d\tau \quad \ldots\ldots \text{(A 12)}$$

As discussed earlier, we are interested only in small values of E, when R_{ik} or $R_{jk} > R_{ij}$. Hence we can approximate (A 16) using

$$1/R_{ij} - 1/R_{ij} = R_{ij}\cos\theta/R^2$$

where R is the mean distance from i and j to k, and θ is the angle between R_{ij} and R. Then

$$P(E) = \frac{1}{2\pi}\int \exp(iE\tau)\exp\{-4N(\pi e^2 R_{ij}/2\kappa)^{3/2}|\tau|^{3/2}\}d\tau,$$

and

$$P(0) = 2\Gamma(5/3)\kappa/\{\pi^2(4N)^{2/3}e^2/R_{ij}\}. \quad \ldots\ldots \text{(A 13)}$$

The appropriate value of W to use in the rectangular distribution (A 13) is therefore

$$W = 1/P(0) = 8{\cdot}5e^2/\kappa a_0 n,$$

where $4\pi N a_0{}^3/3 = 1/n^3$, and n is the average separation in units of the Bohr radius. We have replaced R_{ij} in (A 17) by its average value,

$$R_{ij} = (3/4\pi N)^{1/3}.$$

Hence the critical value of n is determined from (A 12) and (A 15) by

$$(5{\cdot}46 - \ln n)/n^2 \leqslant 0{\cdot}32,$$

and thus on evaluation by $n < n_c$ where

$$n_c = 3{\cdot}6.$$

Our model corresponds to n-type germanium or silicon at the absolute zero of temperature, with degree of compensation $K \simeq 1$. The variation of n_c (the separation of the donor impurities) with K is given by

$$\{5{\cdot}46 - \ln(n/K^{1/3})\}\, n_c{}^2 K^{1/3} = 0{\cdot}32,$$

since the energy spread W is determined by the number of acceptor sites, while the jump distance of the electron depends on the separation of donor sites. Thus n_c increases as K decreases from unity.

A weakness of the above treatment is that V_{ij} and E_{ij} are treated completely independently, whereas from (A 14) and (A 16, 17) it can be seen that both are functions of R_{ij}. Thus we should carry through the treatment using the single quantity V/E as the random variable. Corrections to the

above treatment are being investigated. We believe that the value $n_c = 3{\cdot}6$ is an upper limit to the critical separation. For, we have allowed a finite probability for E_{ij} to have the value zero for all separations R_{ij}. However, for two very close centres i and j there should be a term $2V_{ij}$ in the energy difference E_{ij} which prevents E_{ij} taking a zero value, and therefore the series X (A 7) should converge at smaller values of the average separations between centres.

References

Aigrain, P., 1954, *Physica*, **20**, 978.
Anderson, P. W., 1958, *Phys. Rev.*, **109**, 1492.
Baltensperger, W., 1953, *Phil. Mag.*, **44**, 1355.
Bardeen, J., and Shockley, W., 1950, *Phys. Rev.*, **80**, 72.
Busch, G., and Labhart, H., 1946, *Helv. phys. acta*, **14**, 463.
Carlson, R. O., 1955, *Phys. Rev.*, **100**, 1075.
Chandrasekhar, S., 1943, *Rev. mod. Phys.*, **15**, 1.
Cleland, J. W., Lark-Horovitz, K., and Pigg, J. C., 1950, *Phys. Rev.*, **78**, 814.
Conwell, E. M., 1956, *Phys. Rev.*, **103**, 51.
Dewald, J. F., 1960, *J. Phys. Chem. Solids*, **14**, 155.
Edwards, S. M., 1958, *Phil. Mag.*, **3**, 1020.
Erginsoy, C., 1950, *Phys. Rev.*, **80**, 1104; 1952, *Ibid.*, **88**, 893.
Frisch, H. L., and Lloyd, S. P., 1960, *Phys. Rev.*, **120**, 1175.
Fritzsche, H., 1955, *Phys. Rev.*, **99**, 406; 1958, *J. Phys. Chem. Solids*, **6**, 69; 1960 a, *Phys. Rev.*, **120**, 1120; 1960 b, *Ibid.*, **119**, 1899.
Fritzsche, H., and Cuevas, M., 1960a, *Phys. Rev.*, **119**, 1238; 1960 b, submitted to *Proceedings of the International Conference on Semi-conductor Physics, Prague.*
Fritzsche, H., and Lark-Horovitz, K., 1955, *Phys. Rev.*, **99**, 400; 1959, *Ibid.*, **113**, 999.
Frohlich, H., 1954, *Advance Phys.*, **3**, 325.
Fukoroi, T., Tanuma, S., and Muto, Y., 1954, *Sci. Rep. Res. Inst. Tokuhu Univ.* A, **6**, 18.
Greenwood, D. A., 1958, *Proc. phys. Soc. Lond.*, **71**, 585.
Gudden, B., and Schottky, W., 1935, *Z. tech. Physik*, **16**, 323.
Herring, C., 1959, *J. Phys. Chem. Solids*, **8**, 543.
Hung, C. S., and Gliessman, J. R., 1950, *Phys. Rev.*, **79**, 726; 1954, *Ibid.* **96**, 1226.
Joffe, A. F., 1956, *Canad. J. Phys.*, **34**, 1393.
Kasuya, T., and Koide, S., 1958, *J. phys. Soc., Japan*, **13**, 1287.
Kohn, W., 1957, *Solid State Physics*, Vol. 5 (New York: Academic Press).
Kohn, W., and Luttinger, J. M., 1955, *Phys. Res.* **97**, 883.
Kohn, W., and Schechter, D., 1955, *Phys. Rev.*, **99**, 1903.
Kroger, F. A., Vink, H. J., and Volger, J., 1954, *Physica*, **20**, 1095.
Landau, L., 1933, *Phys. Z. Sowjet.*, **3**, 664.
Lax, M., and Burstein, E., 1955, *Phys. Rev.*, **100**, 592.
Lax, M., and Phillips, J. C., 1958, *Phys. Rev.*, **110**, 41.
Löwdin, P. A., 1956, *Advanc. Phys.*, **5**, 5.
McIrvine, E. C., 1960, *J. Phys. Chem. Solids*, **15**, 356.
Miller, A., and Abrahams, E., 1960, *Phys. Rev.*, **120**, 745.
Morin, F. J., 1958, *Bell Syst. tech., J.*, **37**, 1047.
Morin, F. J., and Maita, J. P., 1954, *Phys. Rev.*, **96**, 28.

Mott, N. F., 1949, *Proc. phys. Soc. Lond.*, **62,** 416; 1952, *Progr. Metal Phys.*, **3,** 76; 1956, *Canad. J. Phys.*, **34,** 1356; 1957, *Nuovo Cim., Suppl.*, **7,** 318; 1961, *Phil. Mag.*, **6,** 287.
Mott, N. F., and Gurney, R. W., 1940, *Electronic Properties of Ionic Crystals.*
Mott, N. F., and Jones, H., 1936, *Theory of the Properties of Metals and Alloys.*
O'Rourke, R. C., 1953, *Phys. Rev.*, **91,** 265.
Price, P. J., 1956, *Phys. Rev.*, **104,** 1223; 1957, *J. Phys. Chem. Solids*, **2,** 268 (Appendix).
Ray, R. K., and Longo, T. A., 1959, *J. Phys. Chem. Solids*, **8,** 259.
Read, P. L., and Katz, E., 1960, *Phys. Rev. Letters*, **5,** 466.
Sewell, G. L., 1958, *Phil. Mag.*, **3,** 1361.
Sladek, R. J., 1956, *J. Phys. Chem. Solids*, **1,** 143; 1958, *Ibid.*, **5,** 157; 1959, *Ibid.*, **8,** 515.
Thomas, D. G., 1959, *J. Phys. Chem. Solids*, **9,** 31.
Twose, W. D., 1959, *Thesis, Cambridge.*
Wigner, E., 1938, *Trans. Faraday Soc.*, **34,** 678.
Yafet, Y., Keyes, R. W., and Adams, E. N., 1956, *J. Phys. Chem. Solids.*, **1,** 137.
Yamashita, J., and Kurosawa, T., 1958, *J. Phys. Chem. Solids*, **5,** 34; 1960, *J. phys. Soc., Japan*, **15,** 802.
Yonemitsu, H., Maeda, H., and Miyazawa, H., 1960, *J. phys. Soc., Japan*, **15,** 1717.
Zaravitskaya, E. I. A., 1956, *J. exp. theor. Phys.*, **30,** 1158.

The General Theory of Van der Waals Forces†

By I. E. Dzyaloshinskii, E. M. Lifshitz and L. P. Pitaevskii
Institute of Physical Problems of the U.S.S.R. Academy of Sciences, Moscow

CONTENTS

§ 1. Introduction

It is well known that there are attractive forces acting between any two neutral atoms or molecules which are separated by a distance large compared to their own dimensions. These are known as van der Waals forces and are of a long-range nature: they decrease with distance according to a power law and not as an exponential.

These van der Waals forces are electromagnetic in origin. As was first shown by London (1930), they arise from second-order perturbation theory applied to the electrostatic interaction between two dipoles; the energy of the interaction being proportional to R^{-6}. However, this approach is only valid so long as R is much less than the wavelengths λ of the corresponding transitions between the ground state and the excited states of the atoms. For $R \gtrsim \lambda$ retardation effects become important. The interaction of atoms taking retardation into account was considered by Casimir and Polder (1948) as an effect of fourth-order perturbation theory applied to the interaction between the atom and the electromagnetic field. (These calculations were repeated by Dzyaloshinskii (1956) using the modern invariant technique of Feynman.) In the limiting case $R \gg \lambda$ the interaction energy is proportional to R^{-7}. The existence of attractive forces between neutral atoms gives rise to analogous forces between any two macroscopic bodies whose surfaces are separated by

† Translated by M. G. Priestley.

very small distances. However, the calculation of these forces on the basis of the interaction between individual atoms (as this is usually done) is impossible. It would be valid only for sufficiently rarefied bodies, i.e. gases—a case which can have no real counterpart. In condensed bodies the close packing of the atoms materially changes the properties of their electronic envelopes, and the presence of some medium between the interacting atoms alters the electromagnetic field through which the interaction is effected.

However, as distinct from this 'microscopic' approach, one may approach the problem from a completely different and purely macroscopic point of view, in which the interacting bodies are considered as continuous media. This approach is valid because the distance between the two surfaces, although small, is large compared to the interatomic distances in the bodies.

The basic idea of the theory is that the interaction between the bodies is considered to take place through a fluctuating electromagnetic field. Because of the thermodynamic fluctuations this field is always present in the interior of a material medium, and it also extends beyond its boundaries. A well-known consequence of this field is the thermal radiation of a body, but it should be emphasized that this is not the only manifestation of the fluctuation field outside the body. This can most clearly be seen already from the fact that electromagnetic fluctuations exist at absolute zero, when there is no thermal radiation; at this temperature the fluctuations are of a purely quantum character.

As well as the attractive forces between bodies placed close together, the same approach can also be used to study other effects in condensed bodies which are due to van der Waals forces, in particular, the properties of thin liquid films on the surface of a solid body.

In a thermodynamical sense, all these effects show one general feature: they are all related to the non-additivity of the free energy of a system of bodies when van der Waals forces are taken into account. In all these cases the free energy is not simply proportional to the volume of the system, but for a given volume is also a function of the parameters which describe the relative positions of the bodies (for example, the distance between solid bodies or the thickness of the film.) It is this non-additivity, due to the long-range nature of van der Waals forces, which is a qualitatively new effect, which separates the contribution of these forces to the thermodynamic quantities from their much larger additive part. This non-additivity can easily be understood by considering the above-mentioned relation between the van der Waals forces and the fluctuations in the electromagnetic field. Any change in the electrical properties of the medium in a certain region will, by Maxwell's equations, lead to a change in the fluctuation field which extends beyond that region. Therefore the part of the free energy which is related to electromagnetic fluctuations is not determined by the properties of the substance solely at the point considered, i.e. it is non-additive.

It should be made clear that when we speak of the fluctuation electromagnetic field, we mean by this all the spectral components which have wavelengths which are large compared to atomic dimensions (we shall refer to these as the long wavelength fluctuations). In each case the important fluctuations are those whose wavelengths are of the same order of magnitude as the inhomogeneities of the system (e.g. the thickness for a film and their separation for the case of two bodies). All the properties of the long wavelength fluctuations, including their contribution to all thermodynamic quantities, are completely specified through the complex dielectric permeability of the body.

In this way it is possible to construct a general macroscopic theory of van der Waals forces which has the sole limitation that all the characteristic dimensions of the bodies must be large compared to interatomic distances. In principle this theory is applicable to any bodies at any temperature, independent of their molecular nature (ionic or molecular crystals, amorphous bodies or liquids, metals, dielectrics, etc.). Since the theory follows from the exact equations of the electromagnetic field, it automatically takes account of retardation effects.

A theory of the attractive van der Waals forces between bodies using these principles was first constructed by Lifshitz (1955). Application of the methods of modern quantum field theory made it possible to find the general formulae for the calculation of the van der Waals part of the thermodynamic quantities for an arbitrary inhomogeneous medium (Dzyaloshinskii and Pitaevskii (1959)). This enables the theory of Lifshitz to be extended to bodies separated by a liquid layer, and also led to its application to the study of the properties of liquid films (Dzyaloshinskii Liftshitz, and Pitaevski (1959)).

We begin the exposition with a short summary of the methods of quantum field theory in statistical physics (§ 2). These methods enable us to develop the whole of the theory of van der Waals forces both naturally and generally. The subsequent exposition is designed so that the reader whose interest lies only in the results of the theory may omit §§2, 3, 4.1.

§ 2. The Methods of Quantum Field Theory in Statistical Physics

The widespread use of the technique of the Feynman diagrams is characteristic of modern quantum field theory ; it enables the structure and the nature of any approximation to be presented very clearly.

As is known, physical quantities in quantum field theory are expressed through perturbation theory series in powers of the interaction constant (for example, in powers of the electronic charge e). Any term of the perturbation theory series can be described by the corresponding diagram and its calculation on the basis of this diagram is governed by the rules of the Feynman technique. In particular, for each internal line of the diagram there corresponds a free particle Green function G_0, or a free photon Green function D_0, for each intersection of lines (a vertex) on the

diagram there is a certain interaction operator (in quantum electrodynamics this is the Dirac matrix γ_μ multiplied by the electronic charge) and, finally, an integration is carried out over the four-dimensional coordinates of each vertex of the diagram.

The advantages of the diagrammatic technique show up most clearly in the solution of problems in which it is not possible to limit the calculations to a finite number of terms of the perturbation series, and it is necessary to sum an infinite series of the so-called 'main diagrams'. The possibility of summing an infinite series makes the diagram technique especially attractive in quantum statistics, the usual methods of which make it very difficult to write down even the first two or three terms of the perturbation theory series.

The application of the methods of quantum field theory to problems in statistical physics at finite temperatures is based on the work of Matsubara (1955), who showed that the free energy could be calculated by the use of the Feynman diagram technique. Each term of the thermodynamic perturbation theory series is, as in field theory, described by the corresponding Feynman diagram and its calculation proceeds along similar lines: each line of the diagram represents the 'temperature' Green function of a free particle $\mathfrak{G}_0$; the vertexes of the diagrams represent the interaction operators. The only difference is that the functions $\mathfrak{G}_0$ in Matsubara's technique depend not on the time t but on the fictitious 'imaginary time' τ which he introduces; this varies over a finite range from 0 to $1/T$, the reciprocal temperature†. Analogously, the time integration from $-\infty$ to $+\infty$ at each diagram intersection is replaced by an integration between 0 and $1/T$.

Here we shall review briefly the work of Matsubara. Consider, for example, a system of charged particles which interact with the electromagnetic field. The Hamiltonian of this system is of the form:

$$H = H_0 + H_{\text{int}}$$

where H_0 is the Hamiltonian of the free particles and photons, and depends quadratically on the corresponding field operators $\psi(\mathbf{r})$ and $A_\alpha(\mathbf{r})$ in the Schrödinger representation, and H_{int} is the interaction operator

$$H_{\text{int}} = -\int A_\alpha(\mathbf{r}) j_\alpha(\mathbf{r})\, d^3\mathbf{r};$$

$j_\alpha(\mathbf{r})$ is the particle current operator‡, which is some quadratic function of the particle operators $\psi(\mathbf{r})$.

The thermodynamical properties of the system are determined by the statistical matrix,

$$\rho = \exp(-H/T),$$

† In §§ 2, 3, 4.1, we use a system of units in which $\hbar = c = 1$; temperature is measured in energy units.

‡ Here and henceforth Greek subscripts α, $\beta = 0$, 1, 2, 3 denote the components of 4-vectors and tensors, while Latin subscripts i, $k = 1$, 2, 3 are used for the components of vectors and tensors in three-dimensional space.

and the free energy F may now be written as

$$F = -T \ln \mathrm{Sp}\rho.$$

In field theory the calculation of the average value of any quantity (here the average value of ρ) necessitates the use of the equations of motion for the field operators. Matsubara's basic idea is the transform from time t to 'imaginary time' τ, which preserves the formal similarity to the usual equations of motion. For this we transform to the appropriate 'interaction representation', which is the analogue of the usual quantum-mechanical interaction representation, using the formulae

$$\begin{aligned}
A_\alpha(\mathbf{r}, \tau) &= \exp(\tau H_0) A_\alpha(\mathbf{r}) \exp(-\tau H_0),\\
\psi(\mathbf{r}, \tau) &= \exp(\tau H_0) \psi(\mathbf{r}) \exp(-\tau H_0),\\
\bar{\psi}(\mathbf{r}, \tau) &= \exp(\tau H_0) \psi^+(\mathbf{r}) \exp(-\tau H_0),\\
j_\alpha(\mathbf{r}, \tau) &= \exp(\tau H_0) j_\alpha(\mathbf{r}) \exp(-\tau H_0),\\
H_{\mathrm{int}}(\tau) &= \exp(\tau H_0) H_{\mathrm{int}} \exp(-\tau H_0);
\end{aligned}$$

It is clear that

$$H_{\mathrm{int}}(\tau) = -\int A_\alpha(\mathbf{r}, \tau) j_\alpha(\mathbf{r}, \tau)\, d^3\mathbf{r}.$$

Next we introduce the matrix

$$\rho(\tau) = \exp(-\tau H)$$

and write it in the form

$$\rho(\tau) = \exp(-\tau H_0) \mathfrak{S}(\tau).$$

The matrix $\mathfrak{S}(\tau)$ is thus defined analogously to the S-matrix in field theory. It satisfies the equation

$$-\frac{\partial \mathfrak{S}(\tau)}{\partial \tau} = H_{\mathrm{int}}(\tau) \mathfrak{S}(\tau), \quad \mathfrak{S}(0) = 1$$

which is obtained from the corresponding equation in field theory by replacing t by $i\tau$. As is known, the solution of this is

$$\mathfrak{S}(\tau) = T_\tau \exp\left\{ -\int_0^\tau H_{\mathrm{int}}(\tau)\, d\tau \right\},$$

where T_τ is the chronologizing operator, which places the operators $H_{\mathrm{int}}(\tau)$ in chronological order for the 'times' τ.

For the statistical matrix $\rho = \rho(1/T)$ we get the obvious formula

$$\rho = \exp(H_0/T) \mathfrak{S},$$

$$\mathfrak{S} \equiv \mathfrak{S}(1/T) = T_\tau \exp\left\{ -\int_0^{1/T} H_{\mathrm{int}}(\tau)\, \mathrm{d}\tau \right\}, \quad \ldots \quad (2.1)$$

whence for the free energy we have

$$F = F_0 - T \ln \mathrm{Sp}\{\exp[(F_0 - H_0)/T] \mathfrak{S}\}, \quad \ldots \quad (2.2)$$

where F_0 is the free energy of the non-interacting particles

$$F_0 = -T \ln \mathrm{Sp} \exp(-H_0/T).$$

Equation (2.2) may be written as

$$F = F_0 - T \ln \langle \mathfrak{S} \rangle, \quad \ldots \quad (2.3)$$

where the symbol $\langle \ldots \rangle$ denotes a Gibbs average over the states of the free particles:

$$\langle \ldots \rangle = \mathrm{Sp}\{\exp[(F_0 - H_0)/T] \ldots\}.$$

When we expand $\mathfrak{S}$ in powers of H_{int}, average each term of the series, and finally take its logarithm, we get a thermodynamical perturbation theory series for the free energy. The above averaging is equivalent to a calculation of the average values of the ordered product of the various electromagnetic field and particle operators, for example

$$\langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\beta(\mathbf{r}_2, \tau_2)\psi(\mathbf{r}_3, \tau_3)\bar{\psi}(\mathbf{r}_4, \tau_4)\}\rangle. \quad \ldots \quad (2.4)$$

Expressions of this type are also met with in quantum field theory.

The technique of Feynman diagrams is based on the following two properties of the equations of field theory: firstly on the possibility of representing all the quantities of the theory (the S-matrix, etc.) as averages of ordered products (T-products) of a different number of field operators, and secondly on the so-called Wick's theorem, according to which the average of the T-product of any number of free particle operators may be expressed in terms of the product of all possible averages of these operators taken in pairs. These pair averages are the free particle Green functions mentioned above. Thus the average of any quantity may be expressed in terms of these Green functions.

Equations (2.1), (2.2) and (2.4) show that the first property is also valid in the thermodynamical theory. In this case Wick's theorem is still valid but becomes here, it is true, an assertion exact only as the total number of particles N tends to infinity (for a given particle density: more accurately it is valid only to terms of order $1/N$.) Applying Wick's theorem to an expression of the type of (2.4) we get, for example,

$$\begin{aligned}
&\langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\beta(\mathbf{r}_2, \tau_2)\psi(\mathbf{r}_3, \tau_3)\bar{\psi}(\mathbf{r}_4, \tau_4)\}\rangle \\
&= \langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\beta(\mathbf{r}_2, \tau_2)\}\rangle\langle T_\tau\{\psi(\mathbf{r}_3, \tau_3)\bar{\psi}(\mathbf{r}_4, \tau_4)\}\rangle, \\
&\langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\beta(\mathbf{r}_2, \tau_2)A_\gamma(\mathbf{r}_3, \tau_3)A_\delta(\mathbf{r}_4, \tau_4)\}\rangle \\
&= \langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\beta(\mathbf{r}_2, \tau_2)\}\rangle\langle T_\tau\{A_\gamma(\mathbf{r}_3, \tau_3)A_\delta(\mathbf{r}_4, \tau_4)\}\rangle \\
&+ \langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\gamma(\mathbf{r}_3, \tau_3)\}\rangle\langle T_\tau\{A_\beta(\mathbf{r}_2, \tau_2)A_\delta(\mathbf{r}_4, \tau_4)\}\rangle \\
&+ \langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\delta(\mathbf{r}_4, \tau_4)\}\rangle\langle T_\tau\{A_\beta(\mathbf{r}_2, \tau_2)A_\gamma(\mathbf{r}_3, \tau_3)\}\rangle.
\end{aligned}$$

It is clear that the technique thus developed is completely analogous to that of field theory, with the sole difference that the zero order Green functions for free particles and photons are replaced by the temperature Green functions

$$\mathfrak{G}^{(0)} = -\langle T_\tau\{\psi(\mathbf{r}_1, \tau_1)\bar{\psi}(\mathbf{r}_2, \tau_2)\}\rangle \quad \ldots \quad (2.4a)$$

for free particles, and

$$\mathfrak{D}_{\alpha\beta}^{(0)} = -\langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\beta(\mathbf{r}_2, \tau_2)\}\rangle \quad \ldots \quad (2.4b)$$

for photons, while the time integration from $-\infty$ to $+\infty$ is replaced by an integration with respect to the 'imaginary time' τ between the limits zero and $1/T$.

The Feynman diagrams which describe corrections to the free energy take the form of closed loops. For the interaction between particles and the electromagnetic field the second-order diagram is given by fig. 1 (*a*), while fig. 1 (*b*), (*c*), (*d*) represent the fourth-order interaction. (Here the full line represents the particle Green function and the dashed line that of the photon.) We note that for a correction of a certain order of perturbation theory we need take into account only the connected diagrams of that order (the order of a diagram is clearly the number of vertexes it contains), i.e. the diagrams which do not separate into components which are not connected at least by one line. For example, fig. 2 need not be taken into account in a sixth-order correction. This property is related to the fact that the expression for the free energy is of the form $\ln \langle \ldots . \rangle$. It can be shown that when the logarithm is taken all the unconnected diagrams cancel out.

Fig. 1

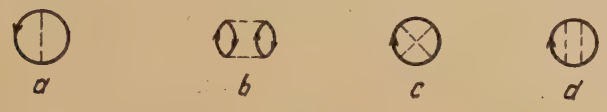


Fig. 2

Fig. 3

The perturbation theory series for the free energy has however one unfortunate shortcoming. It turns out that the diagrams contribute to it with coefficients which vary as $1/n$, where n is the order of the diagram. (Clearly coefficients of the form a^n, where a is a constant, could be dealt with, for example, by the formal inclusion of a in the charge.) This property of the series for the free energy makes it practically worthless for problems where the interaction constant is not small and it is necessary to sum an infinite series of diagrams. Fortunately this applies only to the diagrams which consist of closed loops, for diagrams containing external lines the coefficient does not depend much on the order of the perturbation theory.

The most important of these are those diagrams with two free ends as, for example, the diagrams in fig. 3. The sum of all possible diagrams with two external photon lines is called the full temperature Green function of the photon. It clearly depends on eight variables : spatial coordinates and the

'time' τ of the free ends. It is not difficult to write down an analytical expression for it in terms of the operators in the interaction representation. In particular (cf. (2.4*b*):

$$\mathfrak{D}_{\alpha\beta}(\mathbf{r}_1, \tau_1;\ \mathbf{r}_2, \tau_2) = -\langle T_\tau\{A_\alpha(\mathbf{r}_1, \tau_1)A_\beta(\mathbf{r}_2, \tau_2)\mathfrak{S}\}\rangle/\langle\mathfrak{S}\rangle.$$

Analogous formulae hold for the particle Green functions. We also give a very useful formula, which expresses the full temperature Green function of the photon in terms of the operators in the Schrödinger representation:

$$\mathfrak{D}_{\alpha\beta}(\mathbf{r}_1, \tau_1;\ \mathbf{r}_2, \tau_2) = -\begin{cases} \mathrm{Sp}\{\exp[(F-H)/T]\exp[H(\tau_1-\tau_2)]A_\alpha(\mathbf{r}_1) \\ \quad \exp[-H(\tau_1-\tau_2)]A_\beta(\mathbf{r}_2)\}, \quad \tau_1 > \tau_2, \\ \qquad\qquad\qquad\

The following general property of $\mathfrak{G}$ is fundamental in the manipulation of the perturbation theory series. It follows from the expression (2.5) for $\mathfrak{D}$ that the photon Green function for negative values of τ is related to $\mathfrak{D}$ for $\tau > 0$ by

$$\mathfrak{D}(\tau) = \mathfrak{D}(\tau + 1/T), \quad \tau < 0. \qquad . \; . \; . \; . \; . \qquad (2.7\,a)$$

This also applies to any Bose particle Green function. For Fermi particles we have instead

$$\mathfrak{G}(\tau) = -\mathfrak{G}(\tau + 1/T), \quad \tau < 0. \qquad . \; . \; . \; . \; . \qquad (2.7\,b)$$

Equations (2.7 *a*) and (2.7 *b*) follow easily if we note that we may cyclically permute the orders of the operators under the trace sign in (2.5) and in the analogous formula for Fermi particles. The formulae (2.7 *a*) and (2.7 *b*) are of course also valid for free Green functions.

If we further note than an even number of fermion lines meet at each vertex of the Feynman diagram, it is easy to see that all the integrals of the form $\int_0^{1/T} \ldots d\tau$ in the perturbation theory series may be replaced by $\frac{1}{2}\int_{-1/T}^{1/T} \ldots d\tau$; after this the transformations are easily made. The relations (2.7 *a*) and (2.7 *b*) also mean that the Fourier expansion of boson (or photon) Green functions contains only those components with 'frequencies' $\xi_n = 2\pi n T$, while the expansion of fermions contains only components with $\xi_n = (2n+1)\pi T$.

If we carry out the Fourier transforms with respect to the spatial co-ordinates† and the 'time' τ in all the terms of the perturbation theory series for the Green function (or for the free energy), it is easy to see that the technique which arises thus is fully equivalent to the diagram technique of quantum field theory in the momentum representation. To each line of the diagram there is a corresponding free particle Green function $\mathfrak{G}^{(0)}(\mathbf{p}, \xi_n)$, and for each vertex there is a δ-function which expresses the conservation laws $\sum \mathbf{p} = 0$, $\sum \xi_n = 0$. The integration and summation with respect to all momenta and 'frequencies' are carried out for each line. Formally the expression for the correction due to a certain diagram in the above theory may be obtained from the expression which would correspond to this diagram in field theory by replacing

$$\omega \to i\xi_n; \quad \int_{-\infty}^{\infty} \ldots d\omega \to -2\pi i T \sum_n \ldots$$

The close connection between the above theory and the technique of quantum field theory allows us to apply many results of the latter to the present theory. As in field theory the temperature Green functions satisfy an integral equation of the type of the Dyson equation.

Consider, for example, the diagrams of various orders for the photon Green function. Besides the diagrams in fig. 3, we include those of the type of fig. 4 and others which are more complex.

† This is, of course, only possible for a body which is spatially homogeneous.

All possible diagrams may be represented as shown in fig. 5, where the dashed loops denote the sum of all graphs which do not reduce to components joined by only one photon line. This summation of diagrams is, of course, possible only because the coefficient for a particular diagram does not depend essentially on its order (in the sense used above in the discussion of the series for the free energy).

Thus in order to calculate the total photon Green function it is necessary to sum the series shown schematically in fig. 5. This is of the form (for a spatially inhomogeneous system):

$$\begin{aligned}\mathfrak{D}_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n) = \mathfrak{D}_{\alpha\beta}^{(0)}(\mathbf{r}_1, \mathbf{r}_2; \xi_n) + \int \mathfrak{D}_{\alpha\gamma}^{(0)}(\mathbf{r}_1, \mathbf{r}_3; \xi_n) \\ \times \Pi_{\gamma\delta}(\mathbf{r}_3, \mathbf{r}_4; \xi_n) \mathfrak{D}_{\delta\beta}^{(0)}(\mathbf{r}_4, \mathbf{r}_2; \xi_n)\, d\mathbf{r}_3\, d\mathbf{r}_4 \\ + \int \mathfrak{D}_{\alpha\gamma}^{(0)}(\mathbf{r}_1, \mathbf{r}_3; \xi_n) \Pi_{\gamma\delta}(\mathbf{r}_3, \mathbf{r}_4; \xi_n) \mathfrak{D}_{\delta\mu}^{(0)}(\mathbf{r}_4, \mathbf{r}_5; \xi_n) \\ \times \Pi_{\mu\nu}(\mathbf{r}_5, \mathbf{r}_6; \xi_n) \mathfrak{D}_{\nu\beta}^{(0)}(\mathbf{r}_6, \mathbf{r}_2; \xi_n)\, d\mathbf{r}_3\, d\mathbf{r}_4\, d\mathbf{r}_5\, d\mathbf{r}_6 + \ldots \\ \ldots \qquad (2.8)\end{aligned}$$

Here $\Pi_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n)$ is the so-called polarization operator of the system, which is equal to the sum of the graphs shown in fig. 5 by dashed loops.

Fig. 4

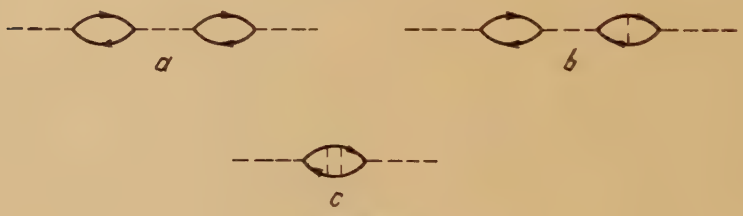


Fig. 5

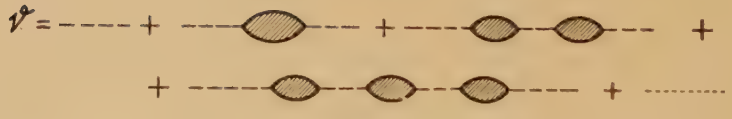


Fig. 6

By re-writing (2.8) as follows

$$\begin{aligned}\mathfrak{D}_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n) = \mathfrak{D}_{\alpha\beta}^{(0)}(\mathbf{r}_1, \mathbf{r}_2; \xi_n) + \int d\mathbf{r}_3\, d\mathbf{r}_4 \mathfrak{D}_{\alpha\gamma}^{(0)}(\mathbf{r}_1, \mathbf{r}_3; \xi_n) \Pi_{\gamma\delta}(\mathbf{r}_3, \mathbf{r}_4; \xi_n) \\ \times \{\mathfrak{D}_{\delta\beta}^{(0)}(\mathbf{r}_4, \mathbf{r}_2; \xi_n) + \int d\mathbf{r}_5\, d\mathbf{r}_6 \cdot \mathfrak{D}_{\delta\mu}^{(0)}(\mathbf{r}_4, \mathbf{r}_5; \xi_n) \Pi_{\mu\nu}(\mathbf{r}_5, \mathbf{r}_6; \xi_n) \\ \times \mathfrak{D}_{\nu\beta}^{(0)}(\mathbf{r}_6, \mathbf{r}_2; \xi_n) + \int d\mathbf{r}_5\, d\mathbf{r}_6\, d\mathbf{r}_7\, d\mathbf{r}_8 \cdot \mathfrak{D}_{\delta\mu}^{(0)}(\mathbf{r}_4, \mathbf{r}_5; \xi_n) \\ \times \Pi_{\mu_\nu}(\mathbf{r}_5, \mathbf{r}_6; \xi_n) \mathfrak{D}_{\nu\lambda}^{(0)}(\mathbf{r}_6, \mathbf{r}_7; \xi_n) \Pi_{\lambda\rho}(\mathbf{r}_7, \mathbf{r}_8; \xi_n) \\ \times \mathfrak{D}_{\rho\beta}^{(0)}(\mathbf{r}_8, \mathbf{r}_9; \xi_n) + \ldots\}\end{aligned}$$

it can easily be seen that it is an integral equation for $\mathfrak{D}$ of the form

$$D_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n) = D_{\alpha\beta}^{(0)}(\mathbf{r}_1, \mathbf{r}_2; \xi_n) + \int \mathfrak{D}_{\alpha\gamma}^{(0)}(\mathbf{r}_1, \mathbf{r}_3; \xi_n) \Pi_{\gamma\delta}(\mathbf{r}_3, \mathbf{r}_4; \xi_n) \mathfrak{D}_{\delta\beta}(\mathbf{r}_4, \mathbf{r}_2; \xi_n) d\mathbf{r}_3 d\mathbf{r}_4. \quad . \quad . \quad . \quad (2.9)$$

The summation process is shown graphically in fig. 6.

In general it is not possible to write a closed equation for the polarization operator, nevertheless the Dyson equation is extremely useful in various practical problems, since it is often possible to find approximate equations for the polarization operator which enable one to go beyond the framework of perturbation theory.

For the long wavelength photons which we shall consider later the polarization operator can be expressed in terms of the dielectric permeability of the body.

§ 3. The Energy of a Condensed Body Associated with Long Wavelength Electromagnetic Fluctuations

We now turn to the solution of our basic problem—the calculation of the increment to the energy of a condensed body due to long wavelength fluctuations of the electromagnetic field. To do this we take out of the Hamiltonian of the whole system that part which represents the interaction energy of particles with the components of the electromagnetic field which have wavelengths much greater than interatomic distances ($\lambda \gg a$), and we shall treat this a perturbation:†

$$H = H_0 + H_{\text{int}} = H_0 - \int A_\alpha(\mathbf{r}) j_\alpha(\mathbf{r}) d\mathbf{r}.$$

The interaction between particles (electrons and nuclei) and the short wavelength field is placed in the unperturbed Hamiltonian. This also includes the short-range interatomic forces which hold the body in the condensed state. The vacuum energy of the long wavelength electromagnetic field is also included in H_0.

We now calculate the corresponding corrections to the free energy. However, as is easily seen, the results of the previous section are not fully applicable to the present case. The Matsubara technique depends on the above-mentioned Wick theorem, according to which the average value of the product of a large number of operators can be expressed as the product of various averages by pairs. However, Wick's theorem is valid only if the Gibbs average is taken over states of non-interacting particles. Here this condition applies only to the operators of the long wavelength electromagnetic field, the averaging of the particle operators being over their states in the condensed body, and therefore the average

† Mathematically the separation of the long wavelength part means that the integral in this equation is in some way cut off for small r. We shall not introduce this cut-off explicitly since the answer does not depend on it.

values of the products of the operators will not reduce to averages by pairs.

We now proceed as follows. In the perturbation series for the free energy (or for the Green function of the long wavelength photons) the particle operators appear only in combinations of the form

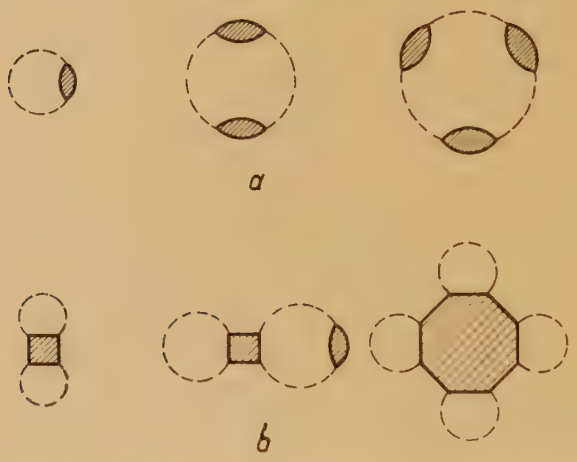
$$\langle T_\tau\{\bar\psi(\mathbf{r}_1,\tau_1)\psi(\mathbf{r}_1,\tau_1)\bar\psi(\mathbf{r}_2,\tau_2)\psi(\mathbf{r}_2,\tau_2)\}\rangle,$$
$$\langle T_\tau\{\bar\psi(\mathbf{r}_1,\tau_1)\psi(\mathbf{r}_1,\tau_1)\bar\psi(\mathbf{r}_2,\tau_2)\psi(\mathbf{r}_2,\tau_2)\bar\psi(\mathbf{r}_3,\tau_3)\psi(\mathbf{r}_3,\tau_3)\bar\psi(\mathbf{r}_4,\tau_4)\psi(\mathbf{r}_4,\tau_4)\}\rangle, \text{ etc.}$$

i.e. the number of operators under the averaging sign is always divisible by four, and the operators always appear in pairs of the type $\bar\psi(\mathbf{r},\tau)\psi(\mathbf{r},\tau)$. We subtract from the average value of the product of eight operators the quantity

$$\langle T_\tau\{\bar\psi(\mathbf{r}_1,\tau_1)\psi(\mathbf{r}_1,\tau_1)\bar\psi(\mathbf{r}_2,\tau_2)\psi(\mathbf{r}_2,\tau_2)\}\rangle\langle T_\tau\{\bar\psi(\mathbf{r}_3,\tau_3)\psi(\mathbf{r}_3,\tau_3)\bar\psi(\mathbf{r}_4,\tau_4)\psi(\mathbf{r}_4,\tau_4)\}\rangle$$
$$+\langle T_\tau\{\bar\psi(\mathbf{r}_1,\tau_1)\psi(\mathbf{r}_1,\tau_1)\bar\psi(\mathbf{r}_3,\tau_3)\psi(\mathbf{r}_3,\tau_3)\}\rangle\langle T_\tau\{\bar\psi(\mathbf{r}_2,\tau_2)\psi(\mathbf{r}_2,\tau_2)\bar\psi(\mathbf{r}_4,\tau_4)\psi(\mathbf{r}_4,\tau_4)\}\rangle$$
$$+\langle T_\tau\{\bar\psi(\mathbf{r}_1,\tau_1)\psi(\mathbf{r}_1,\tau_1)\bar\psi(\mathbf{r}_4,\tau_4)\psi(\mathbf{r}_4,\tau_4)\}\rangle\langle T_\tau\{\bar\psi(\mathbf{r}_2,\tau_2)\psi(\mathbf{r}_2,\tau_2)\bar\psi(\mathbf{r}_3,\tau_3)\psi(\mathbf{r}_3,\tau_3)\}\rangle$$

(i.e. that value which we would have obtained if the averaging had reduced only to all possible averages by fours of the above type), and we call this difference the irreducible quadrilateral, denoting it by a dashed square.

Fig. 7



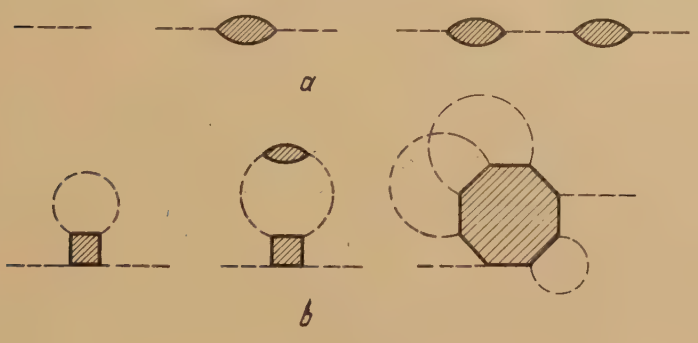
Next, from the average value of twelve operators we subtract the quantity which we would have obtained by reducing it to all possible combinations of four and eight operators. The remainder we call the irreducible hexagon (the dashed hexagons in figs. 7, 8), etc.

It is now not difficult to see that the perturbation theory series will be described by diagrams of the type shown in fig. 7 (for the free energy) and fig. 8 (for the Green function of the long wavelength photons). The dashed loop denotes the quantity obtained from the average value of four

particle operators. The fact that for it we have used a notation which was applied in the previous section to the polarization operator will be justified by later results.

Physically it is immediately clear that the diagrams which contain irreducible quadrilaterals, hexagons, etc. give negligible contributions, since they take into account various non-linear processes such as the scattering of light by light. This assertion can also be proved in another way. Since we have included in H_{int} only the interactions with the long wavelength photons, it should follow that all the integrals over the momenta of the virtual photons are cut off at some k_0 which is considerably smaller than the reciprocal of the interatomic spacing $1/a$. This is evident because each long wavelength photon line over which the integration is taken introduces a small quantity of the order of $k_0 a$. The only diagrams which do not contain an integration over the photon momenta are figs. 7 (a) and 8 (a) (it should be noted that in the zeroth approximation the Green function for the photon depends only on the difference of the coordinates).

Fig. 8



Thus for the approximation $k_0 a \ll 1$ only diagrams of the form 7 (a) give a correction to the free energy. The corresponding expression for the free energy is

$$F = F_0 - \tfrac{1}{2}T \sum_{n=-\infty}^{\infty} \{\Pi_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n)\mathfrak{D}_{\beta\alpha}^{(0)}(\mathbf{r}_2, \mathbf{r}_1; \xi_n)\, d\mathbf{r}_1 d\mathbf{r}_2$$
$$+ \frac{1}{2}\int \Pi_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n)\mathfrak{D}_{\beta\gamma}^{(0)}(\mathbf{r}_2, \mathbf{r}_3; \xi_n)\Pi_{\gamma\delta}(\mathbf{r}_3, \mathbf{r}_4; \xi_n)\mathfrak{D}_{\delta\alpha}^{(0)}(\mathbf{r}_4, \mathbf{r}_1; \xi_n)$$
$$\times d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 d\mathbf{r}_4$$
$$+ \ldots + \frac{1}{m}\int \Pi_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n)\mathfrak{D}_{\beta\gamma}^{(0)}(\mathbf{r}_2, \mathbf{r}_3; \xi_n) \ldots \Pi_{\mu\nu}(\mathbf{r}_{2m-1}\mathbf{r}_{2m}; \xi_n)$$
$$\mathfrak{D}_{\nu\alpha}^{(0)}(\mathbf{r}_{2m}, \mathbf{r}_1; \xi_n)\, d\mathbf{r}_1 \ldots d\mathbf{r}_{2m} + \ldots . \qquad (3.1)$$

where $\Pi_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n)$ is the quantity denoted in the diagram by a dashed loop. Attention is drawn to the coefficients preceding the integrals ($1/m$ for the mth term) which give the contribution from various diagrams.

When, in the same approximation for the photon Green function, we take into account only the contribution from the diagrams of fig 8 (*a*),we obtain, by the same method as was used in the last section in the deduction of the Dyson equation, an equation which is formally identical with (2.9); however, in our approximation the polarization operator Π does not include a contribution from the virtual photon lines and is a given function which depends only on the properties of the body.

The fact that we are dealing with long wavelength photons allows us to express the polarization operator (and also the photon Green function and the free energy of the system) in terms of only the macroscopic parameters of the body. The only quantity which characterizes the interaction between the condensed body and the long wavelength radiation is its dielectric permeability†.

We consider an isotropic (but inhomogeneous) dielectric. The Heisenberg operators for the field strengths satisfy the Maxwell equations (t is ordinary time here)

$$\operatorname{rot}\mathbf{H}=\frac{\partial}{\partial t}(\epsilon\mathbf{E}),\quad \operatorname{rot}\mathbf{E}=-\frac{\partial\mathbf{H}}{\partial t}. \qquad (3.2)$$

In these equations the dielectric permeability is a linear integral operator which acts on functions of t and has the form

$$\epsilon\mathbf{E}(t)=\mathbf{E}(t)+\int_{-\infty}^{t} f(t-t')\mathbf{E}(t')\,dt'. \qquad (3.3)$$

It is difficult to deduce directly from eqns. (3.2) the equations for the operators which depend on the Matsubara 'imaginary time' τ. We therefore use a relation between the temperature Green functions and those of field theory which was obtained by Abrikosov *et al.* (1959) (see also Landau (1958)).

It turns out that the temperature Green function for the photon $\mathfrak{D}_{\alpha\beta}(\mathbf{r}_1,\mathbf{r}_2,\xi_n)$ is simply related to the retarded Green function of the electromagnetic field $D_{\alpha\beta}{}^{\mathrm{R}}$ which is defined as

$$D_{\alpha\beta}{}^{\mathrm{R}}(\mathbf{r}_1,\mathbf{r}_2;\,t_1-t_2)=\begin{cases}-i\mathrm{Sp}\{\exp[(F-H)/T][A_\alpha(\mathbf{r}_1,t_1)A_\beta(\mathbf{r}_2,t_2)-A_\beta(\mathbf{r}_2,t_2)A_\alpha(\mathbf{r}_1,t_1)]\}, & t_1>t_2,\\ 0, & t_1<t_2.\end{cases} \qquad (3.4)$$

(Here the $A_\alpha(\mathbf{r},t)$ are Heisenberg operators). Considerations analogous to those of Abrikosov *et al.* (1959) for a homogeneous body, lead to the conclusion that $\mathfrak{D}_{\alpha\beta}(\mathbf{r}_1,\mathbf{r}_2;\xi_n)$ may be expressed in terms of the Fourier component of the function D^{R}. Explicitly, if we define

$$D_{\alpha\beta}{}^{\mathrm{R}}(\mathbf{r}_1,\mathbf{r}_2;\,\omega)=\int_{-\infty}^{\infty}\exp(i\omega t)D^{\mathrm{R}}(\mathbf{r}_1,\mathbf{r}_2;\,t)\,dt$$

then for $\xi_n>0$ the relation

$$\mathfrak{D}_{\alpha\beta}(\mathbf{r}_1,\mathbf{r}_2;\,\xi_n)=D^{\mathrm{R}}(\mathbf{r}_1,\mathbf{r}_2;\,i\xi_n). \qquad (3.5)$$

† Henceforth we shall neglect the magnetic properties of the substance as they are of no importance in the frequency range considered.

is valid. The value of $\mathfrak{D}_{\alpha\beta}$ for $\xi_n < 0$ can be obtained from the formula for the complex conjugate quantity $\mathfrak{D}_{\alpha\beta}^*$, which follows directly from the definition of the temperature Green function (2.5) and the hermiticity of the electromagnetic field operators:

$$\mathfrak{D}_{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2; \xi_n) = \mathfrak{D}_{\beta\alpha}^*(\mathbf{r}_2, \mathbf{r}_1; -\xi_n). \quad . \quad . \quad . \quad . \quad (3.6)$$

The equation for the retarded function† is now written using the operator Maxwell equations. The question of the gauge of the vector potential now becomes important. The tensor $D_{\alpha\beta}^{\mathrm{R}}$ (or $\mathfrak{D}_{\alpha\beta}$) has ten independent components but we still have at our disposal several arbitary constants because of gauge invariance. Actually physical significance is attached not to the quantities $D^{\mathrm{R}}{}_{\alpha\beta}$ which contain the vector potential components, but only to six corresponding quantities formed from the components of the electric field strength. Thus the ten quantities must satisfy only six physical conditions, i.e. we have four arbitrary functions at our disposal. We may use these to reduce the components D_{00}^{R} and D_{0i}^{R} to zero. Such a choice clearly corresponds to a gauge with a zero scalar potential. The Heisenberg operators $\mathbf{E}$ and $\mathbf{H}$ are now related to $\mathbf{A}$ by the equations

$$\mathbf{E} = \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{H} = \mathrm{rot}\, \mathbf{A},$$

while the operator $\mathbf{A}$ itself satisfies the condition

$$\frac{\partial}{\partial t}\left(\epsilon \frac{\partial \mathbf{A}}{\partial t}\right) + \mathrm{rot}\, \mathrm{rot}\, \mathbf{A} = 0. \quad . \quad . \quad . \quad . \quad . \quad (3.7)$$

For the gauge chosen the commutation relations for the operator $\mathbf{A}$ are of the usual form:

$$\left[\frac{\partial A_i(\mathbf{r}, t)}{\partial t}, A_k(\mathbf{r}', t)\right] = -4\pi i \delta(\mathbf{r} - \mathbf{r}')\delta_{ik}. \quad . \quad . \quad . \quad (3.8)$$

Using eqn. (3.7), the commutation relations (3.8), and transforming to time Fourier components, we get an equation for D_{ik}^{R}:

$$\{\epsilon(\mathbf{r}, \omega)\omega^2\delta_{il} - \mathrm{rot}_{im}\, \mathrm{rot}_{ml}\} D_{lk}^{\mathrm{R}}(\mathbf{r}, \mathbf{r}'; \omega) = 4\pi\delta(\mathbf{r} - \mathbf{r}')\delta_{ik} \quad . \quad . \quad (3.9)$$

Here the symbol rot_{lk} denotes the operator

$$e_{ikl}\, \partial/\partial x_l$$

where e_{ikl} is the alternating tensor. Replacing ω by $i\xi_n$ in (3.9), we find that the function $\mathfrak{D}_{ik}(\mathbf{r}, \mathbf{r}'; \xi_n)$ satisfies the equations

$$\{\epsilon(\mathbf{r}, i\xi_n)\xi_n{}^2\delta_{il} + \mathrm{rot}_{im}\, \mathrm{rot}_{ml}\} \mathfrak{D}_{lk}(\mathbf{r}, \mathbf{r}'; \xi_n) = -4\pi\delta(\mathbf{r} - \mathbf{r}')\delta_{ik} \quad . \quad (3.10)$$

for $\xi_n > 0$.

The dielectric permeability for an imaginary frequency which appears here is related by a simple expression to the imaginary part of the dielectric

† We emphasize that the choice of retarded functions is important, since the operator ϵ relates the value of the field strength at a given moment to its previous values.

permeability for real frequencies (see, for example, Landau and Lifshitz, 1960, §58):

$$\epsilon(i\xi_n)=1+\frac{2}{\pi}\int_0^\infty \frac{\omega\epsilon''(\omega)}{\omega^2+\xi_n^2}d\omega. \quad . \quad . \quad . \quad . \quad (3.11)$$

Since $\epsilon''>0$ it follows that $\epsilon(i\xi_n)$ is a real, positive, and monotonically decreasing function of ξ_n.

Because $\epsilon(i\xi_n)$ is real so is the Green function $\mathfrak{D}_{ik}$ for $\xi_n>0$. For $\xi_n<0$ its value is determined by the relation (see (3.6)):

$$\mathfrak{D}_{ik}(\mathbf{r}_1,\mathbf{r}_2,;\,\xi_n)=\mathfrak{D}_{ki}(\mathbf{r}_2,\mathbf{r}_1;\,-\xi_n) \quad . \quad . \quad . \quad (3.12)$$

By using the Dyson eqn. (2.9), it is not difficult to show that the polarization operator also satisfies the same relation:

$$\Pi_{ik}(\mathbf{r}_1,\mathbf{r}_2;\,\xi_n)=\Pi_{ki}(\mathbf{r}_2,\mathbf{r}_1;\,-\xi_n). \quad . \quad . \quad . \quad . \quad . \quad (3.13)$$

We can now express the polarization operator of the system in terms of $\epsilon(i\xi_n)$. To do this we operate on eqn. (2.9) (for our choice of gauge the components $\mathfrak{D}_{\alpha\beta}$ with $\alpha=0$ or $\beta=0$ are zero) from the left using the operator

$$\xi_n^2\delta_{ik}+\mathrm{rot}_{il}\,\mathrm{rot}_{lk}.$$

Since $\mathfrak{D}$ satisfies (3.10) and $\mathfrak{D}^{(0)}$ satisfies the same equation with $\epsilon(i\xi_n)=1$, we get

$$\int \Pi_{il}(\mathbf{r}_1,\mathbf{r}';\,\xi_n)\mathfrak{D}_{lk}(\mathbf{r}',\mathbf{r}_2;\,\xi_n)\,d\mathbf{r}'=\frac{\epsilon(\mathbf{r}_1,i\xi_n)-1}{4\pi}\xi_n^2\mathfrak{D}_{ik}(\mathbf{r}_1,\mathbf{r}_2;\,\xi_n),$$

whence we see that for $\xi_n>0$

$$\Pi_{ik}(\mathbf{r}_1,\mathbf{r}_2;\,\xi_n)=\frac{\epsilon(\mathbf{r}_1,i\xi_n)-1}{4\pi}\xi_n^2\delta_{ik}\delta(\mathbf{r}_1-\mathbf{r}_2).$$

Defining from (3.13) the polarization operator for $\xi_n<0$, we finally find that for all ξ_n:

$$\Pi_{ik}(\mathbf{r}_1,\mathbf{r}_2;\,\xi_n)=\frac{\epsilon(\mathbf{r}_1,i|\xi_n|)-1}{4\pi}\xi_n^2\delta_{ik}\delta(\mathbf{r}_1-\mathbf{r}_2). \quad . \quad (3.14)$$

The fact that the polarization operator is proportional to $\delta(\mathbf{r}_1-\mathbf{r}_2)$ is related to the neglect of the effects of spatial correlation in the macroscopic theory. These effects are important in metals (particularly in superconductors) at the frequencies of the anomalous skin effect. However we shall be interested in higher frequencies (infra-red and above) in which range there is no spatial dispersion.

Having expressed the polarization operator in terms of the dielectric permeability of the body, we could, in principle, calculate the corresponding correction to the free energy from eqn. (3.1). (The Green function for the free photon can be found directly from the definition (2.4*b*) or by solving eqn. (3.9) with $\epsilon=1$.) However, as we have already pointed out, the series (3.1) cannot be summed directly. Instead we determine the additional pressure (or more accurately an additional stress tensor) which arises as a result of the interaction with the long wavelength fluctuation field.

To do this we imagine that the body is subjected to a certain small displacement described by a vector $\mathbf{u}(\mathbf{r})$. The resultant change in free

energy δF is $-\int \mathbf{f}\mathbf{u}\, dV$, where $\mathbf{f}$ is the force on unit volume of the body due to the deformation. The corresponding change in the unperturbed energy δF_0 is

$$\int \mathbf{u} \operatorname{grad} p_0 \, dV$$

where $p_0(\rho, T)$ is the uncorrected pressure for a given density ρ and temperature T. For a given displacement the only change in the series for the correction is in the polarization operator because only this depends on the properties of the medium. We have

$$\delta\Pi_{ik}(\mathbf{r}_1, \mathbf{r}_2\,;\, \xi_n) = (1/4\pi)\xi_n{}^2\delta_{ik}\delta(\mathbf{r}_1 - \mathbf{r}_2)\delta\epsilon(\mathbf{r}_1, i|\xi_n|).$$

Variation of the series (3.1) gives

$$\begin{aligned}\delta F = \delta F_0 - \frac{T}{8\pi}\sum_{n=-\infty}^{\infty} \xi_n{}^2 \int d\mathbf{r}\,.\,\delta\epsilon(\mathbf{r}_1, i|\xi_n|)\{\mathfrak{D}_{ii}^{(0)}(\mathbf{r}, \mathbf{r}\,;\, \xi_n) \\ + \int \mathfrak{D}_{ik}^{(0)}(\mathbf{r}, \mathbf{r}_1\,;\, \xi_n)\Pi_{kl}(\mathbf{r}_1, \mathbf{r}_2\,;\, \xi_n)\mathfrak{D}_{li}^{(0)}(\mathbf{r}_2, \mathbf{r}\,;\, \xi_n)\, d\mathbf{r}_1 d\mathbf{r}_2 \\ + \int \mathfrak{D}_{ik}^{(0)}(\mathbf{r}, \mathbf{r}_1\,;\, \xi_n)\Pi_{kl}(\mathbf{r}_1, \mathbf{r}_2\,;\, \xi_n)\mathfrak{D}_{lm}^{(0)}(\mathbf{r}_2, \mathbf{r}_3\,;\, \xi_n)\Pi_{mp}(\mathbf{r}_3, \mathbf{r}_4\,;\, \xi_n) \\ \times \mathfrak{D}_{pi}^{(0)}(\mathbf{r}_4, \mathbf{r}\,;\, \xi_n) \quad d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 d\mathbf{r}_4 + \ldots\}.\end{aligned}$$

The series in curly brackets is that for the photon Green function which corresponds to the diagrams of fig. 8 (*a*). Therefore we have

$$\delta F = \delta F_0 - \frac{T}{8\pi}\sum_{n=-\infty}^{\infty} \xi_n{}^2 \int \mathfrak{D}_{ii}(\mathbf{r}, \mathbf{r}\,;\, \xi_n)\delta\epsilon(\mathbf{r}, i|\xi_n|)d\mathbf{r}.$$

By using (3.12) we finally obtain

$$\delta F = \delta F_0 - \frac{T}{4\pi}\sum_{n=0}^{\infty}{}' \xi_n{}^2 \int \mathfrak{D}_{ii}(\mathbf{r}, \mathbf{r}\,;\, \xi_n)\delta\epsilon(\mathbf{r}, i\xi_n)\, d\mathbf{r} \quad . \quad . \quad (3.15)$$

The dash on the summation sign denotes that the term with $n = 0$ is given half weight. We note that $\xi_n = 2\pi nT$.

The variation of ϵ is related to the displacement $\mathbf{u}$ by the equation

$$\delta\epsilon = -\mathbf{u} \operatorname{grad} \epsilon - \rho\frac{\partial\epsilon}{\partial\rho}\operatorname{div}\mathbf{u}. \quad . \quad . \quad . \quad . \quad . \quad (3.16)$$

Substituting this in (3.15) and integrating by parts we obtain an expression for the force acting on unit volume of the body

$$\begin{aligned}\mathbf{f} = -\operatorname{grad} p_0 - \frac{T}{4\pi}\sum_{n=0}^{\infty}{}' \xi_n{}^2\mathfrak{D}_{ii}(\mathbf{r}, \mathbf{r}\,;\, \xi_n)\operatorname{grad}\epsilon \\ + \frac{T}{4\pi}\sum_{n=0}^{\infty}{}' \xi_n{}^2 \operatorname{grad}\left\{\mathfrak{D}_{ii}(\mathbf{r}, \mathbf{r}\,;\, \xi_n)\rho\frac{\partial\epsilon}{\partial\rho}\right\}. \quad . \quad . \quad (3.17)\end{aligned}$$

From this formula it is easy to calculate the correction to the chemical potential of the body. First we note that $\mathbf{f}=0$ in mechanical equilibrium and that for a given temperature the relations

$$\operatorname{grad}\epsilon(\rho, T)=\frac{\partial\epsilon}{\partial\rho}\operatorname{grad}\rho, \quad dp_0(\rho, T)=\rho d\zeta_0(\rho, T) \qquad (3.18)$$

are valid (where $\zeta_0(\rho, T)$ is the unperturbed chemical potential per unit mass of the body), then we equate (3.17) to zero and a simple transformation gives

$$\rho\operatorname{grad}\zeta=0,$$

where

$$\zeta(\rho, T)=\zeta_0(\rho, T)-\frac{T}{4\pi}\sum_{n=0}^{\infty}{}'\,\xi_n{}^2\mathfrak{D}_{ii}(\mathbf{r}, \mathbf{r}; \xi_n)\frac{\partial\epsilon}{\partial\rho}. \qquad (3.19)$$

As is known, the equilibrium condition for any inhomogeneous body is that the chemical potential should be constant over it; it is therefore clear that (3.19) defines this potential (per unit mass).

We now turn to the calculation of the pressure. We need to reduce the expression (3.17) for the force per unit volume of the body to the form

$$f_i=\frac{\partial\sigma_{ik}}{\partial x_k} \qquad (3.20)$$

where σ_{ik} is the strain tensor. The resultant calculations are almost identical with those carried out in electrodynamics to find the Maxwell stress tensor (see, for example, Landau and Lifshitz 1960, §15), but we shall discuss them briefly here.

First we introduce, besides the photon Green function $\mathfrak{D}_{ik}$, two related functions

$$\left.\begin{aligned}\mathfrak{D}_{ik}{}^{E}(\mathbf{r}, \mathbf{r}'; \xi_n)&=-\xi_n{}^2\mathfrak{D}_{ik}(\mathbf{r}, \mathbf{r}'; \xi_n),\\ \mathfrak{D}_{ik}{}^{E}(\mathbf{r}, \mathbf{r}'; \xi_n)&=\operatorname{rot}_{il}\operatorname{rot}'_{km}\mathfrak{D}_{lm}(\mathbf{r}, \mathbf{r}'; \xi_n),\end{aligned}\right\} \qquad (3.21)$$

which are made up from the electric and magnetic field operators in the same way that $\mathfrak{D}_{ik}$ is made up from the operators of the vector potential.

We now rewrite (3.17) in the new notation†

$$f_i=-\frac{\partial p_0}{\partial x_i}+\frac{T}{4\pi}\sum{}'\frac{\partial}{\partial x_i}\left\{\epsilon(\mathbf{r})\mathfrak{D}_{kk}{}^{E}(\mathbf{r}, \mathbf{r})-\rho\frac{\partial\epsilon}{\partial\rho}\mathfrak{D}_{kk}{}^{E}(\mathbf{r}, \mathbf{r})\right\}$$
$$-\frac{T}{4\pi}\sum{}'\epsilon(\mathbf{r})\frac{\partial}{\partial x_i}\mathfrak{D}_{kk}{}^{E}(\mathbf{r}, \mathbf{r}). \qquad (3.22)$$

We now transform just the last term in (3.22); taking out the summation and the factor $T/4\pi$, we rewrite it in the form

$$\epsilon(\mathbf{r}')\frac{\partial}{\partial x_i}\mathfrak{D}_{kk}{}^{E}(\mathbf{r}, \mathbf{r}')+\epsilon(\mathbf{r})\frac{\partial}{\partial x_i'}\mathfrak{D}_{kk}{}^{E}(\mathbf{r}, \mathbf{r}')$$

with the intention of putting $\mathbf{r}=\mathbf{r}'$ at the end of the calculations.

† To be concise, we shall omit the arguments ξ_n and $i\xi_n$ in the intermediate formulae.

Carrying out further obvious transformations we get

$$2\frac{\partial}{\partial x_k}\epsilon(\mathbf{r})\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r}')-\frac{\partial}{\partial x_k}\epsilon(\mathbf{r})\mathfrak{D}_{ki}^{E}(\mathbf{r},\mathbf{r}')-\frac{\partial}{\partial x_k'}\epsilon(\mathbf{r}')\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r}')$$
$$+\epsilon(\mathbf{r}')\left[\frac{\partial}{\partial x_i}\mathfrak{D}_{kk}^{E}(\mathbf{r},\mathbf{r}')-\frac{\partial}{\partial x_k}\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r}')\right]$$
$$+\epsilon(\mathbf{r})\left[\frac{\partial}{\partial x_i'}\mathfrak{D}_{kk}^{E}(\mathbf{r},\mathbf{r}')-\frac{\partial}{\partial x_k'}\mathfrak{D}_{ki}^{E}(\mathbf{r},\mathbf{r}')\right] \quad (3.23)$$

The equation for the Green function (3.10) gives the following identities:

$$\frac{\partial}{\partial x_k}\epsilon(\mathbf{r})\mathfrak{D}_{ki}^{E}(\mathbf{r},\mathbf{r}')=4\pi\frac{\partial}{\partial x_i}\delta(\mathbf{r}-\mathbf{r}'),$$
$$\frac{\partial}{\partial x_k'}\epsilon(\mathbf{r}')\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r}')=-4\pi\frac{\partial}{\partial x_i}\delta(\mathbf{r}-\mathbf{r}'),$$
$$\epsilon(\mathbf{r}')\left[\frac{\partial}{\partial x_k}\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r}')-\frac{\partial}{\partial x_i}\mathfrak{D}_{kk}^{E}(\mathbf{r},\mathbf{r}')\right]=$$
$$-\frac{\partial}{\partial x_k'}\mathfrak{D}_{ki}^{H}(\mathbf{r},\mathbf{r}')+\frac{\partial}{\partial x_i'}\mathfrak{D}_{kk}^{H}(\mathbf{r},\mathbf{r}')-8\pi\frac{\partial}{\partial x_i}\delta(\mathbf{r}-\mathbf{r}'),$$
$$\epsilon(\mathbf{r})\left[\frac{\partial}{\partial x_k'}\mathfrak{D}_{ki}^{E}(\mathbf{r},\mathbf{r}')-\frac{\partial}{\partial x_i'}\mathfrak{D}_{kk}^{E}(\mathbf{r},\mathbf{r}')\right]=$$
$$-\frac{\partial}{\partial x_k}\mathfrak{D}_{ik}^{H}(\mathbf{r},\mathbf{r}')+\frac{\partial}{\partial x_i}\mathfrak{D}_{kk}^{H}(\mathbf{r},\mathbf{r}')+8\pi\frac{\partial}{\partial x_i}\delta(\mathbf{r}-\mathbf{r}').$$

Substituting these in (3.21) and putting $\mathbf{r}=\mathbf{r}'$ gives

$$\epsilon(\mathbf{r})\frac{\partial}{\partial x_i}\mathfrak{D}_{kk}^{E}(\mathbf{r},\mathbf{r})=2\frac{\partial}{\partial x_k}\epsilon(\mathbf{r})\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r})+2\frac{\partial}{\partial x_k}\mathfrak{D}_{ik}^{H}(\mathbf{r},\mathbf{r})-\frac{\partial}{\partial x_i}\mathfrak{D}_{kk}^{H}(\mathbf{r},\mathbf{r}).$$

Finally substitution in (3.22) shows that the force can be expressed in the form (3.20) with a stress tensor

$$\sigma_{ik}=-p_0(\rho,T)\delta_{ik}-\frac{T}{2\pi}\sum_{n=0}^{\infty}{}'\left\{-\tfrac{1}{2}\delta_{ik}\left[\epsilon(\mathbf{r};i\xi_n)-\rho\frac{\partial\epsilon(\mathbf{r},i\xi_n)}{\partial\rho}\right]\mathfrak{D}_{ll}^{E}(\mathbf{r},\mathbf{r};\xi_n)\right.$$
$$\left.+\epsilon(\mathbf{r},i\xi_n)\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r};\xi_n)-\tfrac{1}{2}\delta_{ik}\mathfrak{D}_{ll}^{H}(\mathbf{r},\mathbf{r};\xi_n)+\mathfrak{D}_{ik}^{H}(\mathbf{r},\mathbf{r};\xi_n)\right\}. \quad (3.24)$$

However, this formula still does not have a direct physical meaning, since the quantities $\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r}')$ and $\mathfrak{D}_{ik}^{H}(\mathbf{r},\mathbf{r}')$ both tend to infinity for $\mathbf{r}=\mathbf{r}'$. This is related to the fact that, unless a corresponding cut-off is introduced, fluctuations with small wavelengths make an infinitely large contribution to σ_{ik}. These have no relation to the inhomogeneity of the body in the sense that their contribution is the same in both homogeneous and inhomogeneous bodies which have the same value of ϵ at the point considered. The contribution of the long wavelength fluctuations to the stress tensor in an inhomogeneous medium, which is actually independent of the nature of the cut-off, is obtained by the corresponding subtraction in (3.24). The Green function $\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r}')$ (and analogously for $\mathfrak{D}_{ik}^{H}$) in this formula must be interpreted as the limit of the difference

$$\lim_{\mathbf{r}'\to\mathbf{r}}[\mathfrak{D}_{ik}^{E}(\mathbf{r},\mathbf{r}')-\overline{\mathfrak{D}}_{ik}^{E}(\mathbf{r},\mathbf{r}')]$$

where $\overline{\mathfrak{D}}_{ik}{}^{E}$ is the Green function of an homogeneous infinite medium whose dielectric permeability coincides with that for the inhomogeneous medium at the point where the stress tensor is being calculated. To avoid excessively unwieldy formulae we shall in future write (3.24) in its previous form with the assumption that the above subtraction has already been carried out. Here $p_0(\rho, T)$ is the pressure in an infinite homogeneous medium for given values of ρ and T.

The same remarks also apply to formula (3.19) for the chemical potential, which by using (3.21) can be written in the form

$$\zeta(\rho, T) = \zeta_0(\rho, T) + \frac{T}{4\pi} \sum_{n=0}^{\infty}{}' \frac{\partial \epsilon(\mathbf{r}\,;\, i\xi_n)}{\partial \rho} \mathfrak{D}_{ii}{}^{E}(\mathbf{r}, \mathbf{r}\,;\, \xi_n). \qquad (3.25)$$

We note that the category of inhomogeneous media includes systems which consist of several bodies, each of which is homogeneous. In this case the components $\mathfrak{D}_{ik}$ in the solution of the eqns. (3.10) must satisfy certain conditions at the boundaries between the bodies. We recall that in (3.10) the coordinates $\mathbf{r}$ are independent variables, while the coordinates $\mathbf{r}'$ are merely parameters. Hence we discuss the boundary conditions for the variables $\mathbf{r}$. These conditions correspond to the continuity of the tangential components of the electric and magnetic fields. Since one of the indices (i) of the tensor $\mathfrak{D}_{ik}$ refers to the point $\mathbf{r}$ (in the sense of the definitions (3.4)), the tangential components of the tensors $\mathfrak{D}_{ik}{}^{E}$ and $\mathfrak{D}_{ik}{}^{H}$ must be continuous with respect to this index.

In form eqn. (3.24) is similar to the usual formula for the Maxwell stresses in an electromagnetic field, but the quadratic combinations of components of the electric and magnetic fields are replaced by the corresponding functions $\mathfrak{D}_{ik}{}^{E}$ and $\mathfrak{D}_{ik}{}^{H}$. Too deep a significance should not, however, be attached to this analogy. The fact is that there are good grounds for the supposition that the concept of the stress tensor for an alternating electromagnetic field in a conducting medium is not meaningful. But in (3.24) we are dealing not with an arbitrary electromagnetic field but with the proper fluctuation field which is in thermodynamical equilibrium with the medium.

Equations (3.24) and (3.25), obtained by Dzyaloshinskii and Pitaevskii (1959), solve in principle the problem of the calculation of the van der Waals part of the thermodynamic parameters of a body, by reducing it in each real case to the solution of the eqns. (3.10) for the Green function $\mathfrak{D}_{ik}$.

§ 4. Molecular Interaction Forces between Solid Bodies

4.1. *Derivation of General Formulae*

We use the general theory developed above to calculate the van der Waals forces acting between solid bodies whose surfaces are separated by very small distances. The gap separating the bodies may be considered to be filled with any liquid. In what follows we shall denote the two solid bodies by the indices 1 and 2, while the medium which fills the gap will be denoted by the index 3.

Although we shall represent the gap by two parallel planes, it should be noted that for a rigorous statement of the problem it would be necessary to consider at least one of the bodies as being of finite size and as surrounded by the medium 3 on all sides, the total forces on the body then being calculated. However, since molecular forces die away very quickly with distance this resultant force is in fact completely equivalent to the forces acting across the narrow gap which separates the two bodies.

The total force acting on the body 2 can be calculated as the total momentum flux which enters the body from the surrounding medium 3, i.e. as an integral $\int \sigma_{ik}\, df_k$ over the surface surrounding it. For this it should be noted that the medium 3 is in thermodynamical equilibrium, one of the conditions of which is that its chemical potential should be constant: $\zeta = \text{const}$, where ζ is given by eqn. (3.25). Since the long wavelength fluctuations give only small corrections to the density, we can take the density ρ to be constant throughout medium 3, and the change in the chemical potential $\zeta_0(\rho, T)$ equals the change in the quantity $\rho_0(\rho, T)/\rho$ (because (3.18) applies). Therefore the condition $\zeta = \text{const}$ can be written as:

$$p_0(\rho, T) + \frac{T}{4\pi} \sum{}' \rho \frac{\partial \epsilon_3}{\partial \rho} \mathfrak{D}_{ll}{}^{E}(\mathbf{r}, \mathbf{r}) = \text{const.} \qquad . \; . \; . \quad (4.1)$$

As a consequence of this condition part of the total stress tensor is a constant equilibrium pressure throughout the liquid, and this does not contribute to the total force acting on the body. Neglecting this constant part, i.e. subtracting from σ_{ik} the left-hand side of (4.1) multiplied by δ_{ik}, we find that to determine the force it is enough to write the stress tensor in medium 3 in the form

$$\sigma_{ik}{}' = -\frac{T}{2\pi} \sum{}' \{\epsilon_3[\mathfrak{D}_{ik}{}^{E}(\mathbf{r}, \mathbf{r}) - \tfrac{1}{2}\delta_{ik}\mathfrak{D}_{ll}{}^{E}(\mathbf{r}, \mathbf{r})] + \mathfrak{D}_{ik}{}^{H}(\mathbf{r}, \mathbf{r}) - \tfrac{1}{2}\delta_{ik}\mathfrak{D}_{ll}{}^{H}(\mathbf{r}, \mathbf{r})\}.$$
$$. \; . \; . \; . \; . \; . \; . \quad (4.2)$$

We take the x-axis perpendicular to the plane of the gap, whose breadth is denoted by l (so the surfaces of the bodies are given by the planes $x = 0$ and $x = l$). We can now write the force per unit area of body 2 as

$$F(l) = \sigma_{xx}{}'(l) = \frac{T}{4\pi} \sum_{n=0}^{\infty}{}' \{\epsilon_3[\mathfrak{D}_{yy}{}^{E}(l, l\,;\, \xi_n) + \mathfrak{D}_{zz}(l, l\,;\, \xi_n) - \mathfrak{D}_{xx}{}^{E}(l, l\,;\, \xi_n)]$$
$$+ \mathfrak{D}_{yy}{}^{H}(l, l\,;\, \xi_n) + \mathfrak{D}_{zz}{}^{H}(l, l\,;\, \xi_n) - \mathfrak{D}_{xx}{}^{H}(l, l\,;\, \xi_n)\}; \qquad . \; . \quad (4.3)$$

where a positive force corresponds to an attraction and a negative to repulsion.

Because the problem is homogeneous in y and z the Green function $\mathfrak{D}_{ik}(\mathbf{r}, \mathbf{r}')$ is a function only of $y - y'$ and $z - z'$. We make a Fourier transform with respect to these variables

$$\mathfrak{D}_{ik}(x, x'\,;\, \mathbf{q}\,;\, \xi_n) = \int\int \exp\left[-iq_y(y - y') - iq_z(z - z')\right]$$
$$\times \mathfrak{D}_{ik}(\mathbf{r}, \mathbf{r}'\,;\, \xi_n)\, d(y - y')\, d(z - z')$$

and take the y-axis along the vector $\mathbf{q}$. The eqns. (3.10) for the Green functions now become

$$\left(w^2-\frac{d^2}{dx^2}\right)\mathfrak{D}_{zz}(x,x')=-4\pi\delta(x-x'),$$

$$\left(w^2-q^2-\frac{d^2}{dx^2}\right)\mathfrak{D}_{yy}(x,x')+iq\frac{d}{dx}\mathfrak{D}_{xy}(x,x')=-4\pi\delta(x-x'),$$

$$w^2\mathfrak{D}_{xy}(x,x')+iq\frac{d}{dx}\mathfrak{D}_{yy}(x,x')=0,$$

$$w^2\mathfrak{D}_{xx}(x,x')+iq\frac{d}{dx}\mathfrak{D}_{xy}(x,x')=-4\pi\delta(x-x'),$$

$$\left(w^2-q^2-\frac{d^2}{dx^2}\right)\mathfrak{D}_{xy}(x,x')+iq\frac{d}{dx}\mathfrak{D}_{xx}(x,x')=0,$$

where $w=(\epsilon\xi_n{}^2+q^2)^{1/2}$, and x' is a parameter (the components $\mathfrak{D}_{yz}$ and $\mathfrak{D}_{xz}$ of the Green function are zero because the equations for them are homogeneous).

The solution of this system reduces to the solution of two equations

$$\left.\begin{aligned}\left(w^2-\frac{d^2}{dx^2}\right)\mathfrak{D}_{zz}(x,x')&=-4\pi\delta(x-x'),\\ \left(w^2-\frac{d^2}{dx^2}\right)\mathfrak{D}_{yy}(x,x')&=-\frac{4\pi w^2}{\epsilon\xi_n{}^2}\delta(x-x'),\end{aligned}\right\}\quad\ldots\quad(4.4)$$

after which $\mathfrak{D}_{xy}$ and $\mathfrak{D}_{xx}$ are calculated as

$$\left.\begin{aligned}\mathfrak{D}_{xy}(x,x')&=-\frac{iq}{w^2}\frac{d}{dx}\mathfrak{D}_{yy}(x,x'),\\ \mathfrak{D}_{xx}(x,x')&=-\frac{iq}{w^2}\frac{d}{dx}\mathfrak{D}_{xy}(x,x')-\frac{4\pi}{w^2}\delta(x-x').\end{aligned}\right\}\quad\ldots\quad(4.5)$$

The boundary conditions which correspond to the continuity of the tangential components of the electric and magnetic fields reduce to the requirement that the quantities $\mathfrak{D}_{yk}{}^E$, $\mathfrak{D}_{yk}{}^H$, $\mathfrak{D}_{zk}{}^E$, $\mathfrak{D}_{zk}{}^H$ should be continuous, or equivalently that

$$\mathfrak{D}_{yk},\ \mathfrak{D}_{zk},\ \mathrm{rot}_{yl}\mathfrak{D}_{lk},\ \mathrm{rot}_{zl}\mathfrak{D}_{lk}$$

should be continuous.

Using (4.5), we find that

$$\mathfrak{D}_{zz},\ \frac{d}{dx}\mathfrak{D}_{zz},\ \mathfrak{D}_{yy},\ \frac{\epsilon}{w^2}\frac{d}{dx}\mathfrak{D}_{yy}\quad\ldots\quad(4.6)$$

must be continuous at the interface.

Since we are interested in the Green function only in the gap we can straightaway limit discussion to $0<x'<l$. In the range $0<x<l$ the functions $\mathfrak{D}_{yy}$ and $\mathfrak{D}_{zz}$ are given by eqns. (4.4) with $\epsilon=\epsilon_3$, $w=w_3=(\epsilon_3\xi_n{}^2+q^2)^{1/2}$. In the regions 1 ($x<0$) and 2 ($x>l$) they satisfy the same equations without the right-hand sides (because here $x\neq x'$ always) with ϵ, w replaced by ϵ_1, w_1 and ϵ_2, w_2 respectively.

The subtraction mentioned at the end of § 3 means here that from all the functions $\mathfrak{D}_{ik}$ in the region of the gap we must subtract their values for $\epsilon_1=\epsilon_2=\epsilon_3$, $w_1=w_2=w_3$. Hence, in particular, we can drop the term containing the δ-functions in the second of the equations (4.5), so that in the gap the functions $\mathfrak{D}_{xy}$, $\mathfrak{D}_{xx}$ are given by the equations

$$\mathfrak{D}_{xy}=-\frac{iq}{{w_3}^2}\frac{d\mathfrak{D}_{yy}}{dx},\ \mathfrak{D}_{xx}=-\frac{iq}{{w_3}^2}\frac{d\mathfrak{D}_{xy}}{dx}. \quad . \quad . \quad . \quad (4.7)$$

Before we solve these equations we make one further remark. The general solution of the eqn. (4.4) is of the form $f^+(x-x')+f^-(x+x')$. By using (4.4), (4.7) and the definition of the functions $\mathfrak{D}_{ik}{}^E$ and $\mathfrak{D}_{ik}{}^H$ it can be shown that the parts of the Green functions which depend on the sum $x+x'$ make no contribution to the expression (4.3) for the force F. We shall not discuss this further here since it is already obvious from physical considerations: if we put $x=x'$ in a solution of the form $f^-(x+x')$ we would obtain a momentum flux in the gap which varied with the coordinates, and this would contradict its conservation. Therefore, below we shall give only the expression for the part of the Green function which depends on $x-x'$.

We now turn to the determination of the function $\mathfrak{D}_{zz}$. This satisfies the equations

$$\left({w_3}^2-\frac{d^2}{dx^2}\right)\mathfrak{D}_{zz}(x,x')=-4\pi\delta(x-x'), \qquad 0<x<l,$$

$$\left({w_1}^2-\frac{d^2}{dx^2}\right)\mathfrak{D}_{zz}(x,x')=0, \qquad x<0,$$

$$\left({w_2}^2-\frac{d^2}{dx^2}\right)\mathfrak{D}_{zz}(x,x')=0, \qquad x>l.$$

Hence we find

$$\mathfrak{D}_{zz}=A\exp(w_1x), \qquad x<0,$$

$$\mathfrak{D}_{zz}=B\exp(-w_2x), \qquad x>l,$$

$$\mathfrak{D}_{zz}=C_1\exp(w_3x)+C_2\exp(-w_3x)-\frac{2\pi}{w_3}\exp(-w_3|x-x'|), \qquad 0<x<l.$$

Determining the constants A, B, C_1, C_2 from the boundary conditions that $\mathfrak{D}_{zz}$ and $\mathfrak{D}_{zz}/dx$ are continuous, we find that $\mathfrak{D}_{zz}{}^+$

$$\mathfrak{D}_{zz}{}^+=\frac{4\pi}{w_3\Delta}\operatorname{ch}w_3(x-x')-\frac{2\pi}{w_3}\exp(-w_3|x-x'|), \qquad 0<x<l,$$

where

$$\Delta=1-\exp(2w_3l)\frac{(w_1+w_3)(w_2+w_3)}{(w_1-w_3)(w_2-w_3)}. \quad . \quad . \quad . \quad (4.8)$$

Subtracting the value of $\mathfrak{D}_{zz}{}^+$ for $w_1=w_2=w_3$ (when $1/\Delta$ becomes zero), we finally obtain

$$\mathfrak{D}_{zz}{}^+=\frac{4\pi}{w_3\Delta}\operatorname{ch}w_3(x-x'). \quad . \quad . \quad . \quad . \quad . \quad (4.9)$$

Analogously, solution for $\mathfrak{D}_{yy}$ gives (after subtraction)

$$\mathfrak{D}_{yy}{}^{+}=\frac{4\pi w_3}{\xi_n{}^2\epsilon_3\bar{\Delta}}\operatorname{ch} w_3(x-x'), \quad . \quad . \quad . \quad . \quad . \quad (4.10)$$

$$\bar{\Delta}=1-\exp(2w_3 l)\frac{(\epsilon_1 w_3+\epsilon_3 w_1)(\epsilon_2 w_3+\epsilon_3 w_2)}{(\epsilon_1 w_3-\epsilon_3 w_1)(\epsilon_2 w_3-\epsilon_3 w_2)} \quad . \quad . \quad (4.11)$$

and, by using (4.7)

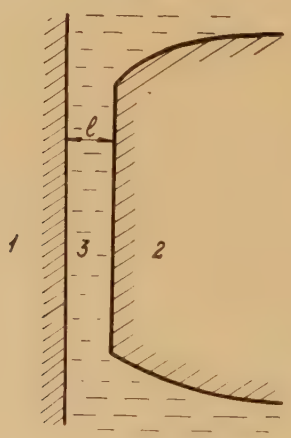
$$\left.\begin{aligned}\mathfrak{D}_{xy}{}^{+}&=-\frac{4\pi i q}{\xi_n{}^2\epsilon_3\bar{\Delta}}\operatorname{sh} w_3(x-x'),\\ \mathfrak{D}_{xx}{}^{+}&=-\frac{4\pi q^2}{\xi_n{}^2\epsilon_3 w_3\bar{\Delta}}\operatorname{ch} w_3(x-x').\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (4.12)$$

If we now calculate the quantities $\mathfrak{D}_{ik}(x, x'; \mathbf{q}; \xi_n)$ and $\mathfrak{D}_{ik}{}^{H}(x, x; \mathbf{q}; \xi_n)$ and substitute them in eqn. (4.3), we get:

$$F(l)=-\frac{T}{2\pi}\sum_{n=0}{}'^{\infty}\int_0^\infty q\,dq\,.\,w_3\left(\frac{1}{\Delta}+\frac{1}{\bar{\Delta}}\right).$$

Transforming to a new variable of integration p, where $q=\xi_n[(\epsilon_3(p^2-1)]^{1/2}$ and returning to a normal system of units, we obtain the final expression for the force F acting on unit area of each of the two bodies (media 1 and 2) separated by a gap of width l which is occupied by medium 3 (fig. 9)

Fig. 9



$$F(l)=\frac{kT}{\pi c^3}\sum_{n=0}{}'^{\infty}\epsilon_3{}^{3/2}\xi_n{}^3\int_1^\infty p^2\left\{\left[\frac{(s_1+p)(s_2+p)}{(s_1-p)(s_2-p)}\exp\left(\frac{2p\xi_n}{c}l\sqrt{\epsilon_3}\right)-1\right]^{-1}\right.$$
$$\left.+\left[\frac{(s_1+p\epsilon_1/\epsilon_3)(s_2+p\epsilon_2/\epsilon_3)}{(s_1-p\epsilon_1/\epsilon_3)(s_2-p\epsilon_2/\epsilon_3)}\exp\left(\frac{2p\xi_n}{c}l\sqrt{\epsilon_3}\right)-1\right]^{-1}\right\}dp \quad (4.13)$$

where $s_1=\sqrt{(\epsilon_1/\epsilon_3-1+p^2)}$, $s_2=\sqrt{(\epsilon_2/\epsilon_3-1+p^2)}$, $\xi_n=2\pi nkT/h$; ϵ_1, ϵ_2, ϵ_3 are functions of the imaginary frequency $\omega=i\xi_n$ $(\epsilon=\epsilon(i\xi_n))$; k is Boltzmann's constant. The summation is taken over integral values of n,

and the dash attached to the summation sign denotes that the term with $n=0$ is given half weight. Positive values of F correspond to attraction between the bodies and negative values to repulsion.

This formula (for $\epsilon_3=1$, i.e. for bodies separated by an empty gap) was first derived by Lifshitz (1954) by another method which did not use the techniques of quantum field theory. These techniques are however necessary for the generalization of the result to a gap filled by an arbitrary medium (Dzyaloshinskii *et al.* 1959).

4.2. *Discussion of General Formulae and Limiting Cases*†

The general formula (4.13) contains the functions $\epsilon(\omega)$—the dielectric permeabilities as a function of field frequency—for both solid bodies (ϵ_1 and ϵ_2) and for the fluid medium which fills the space between them (ϵ_3). We note that $\epsilon(\omega)$ is a complex quantity $[\epsilon=\epsilon'(\omega)+i\epsilon''(\omega)]$, and that its imaginary part is always positive and determines the dissipation of energy in an electromagnetic wave propagated in the medium. The function $\epsilon(\omega)$ is related to the refractive index n and the absorption coefficient κ by the well-known expression $\sqrt{\epsilon}=n+i\kappa$. It is known that a formal consideration of $\epsilon(\omega)$ as a function of the complex variable ω leads to certain integral relations between $\epsilon'(\omega)$ and $\epsilon''(\omega)$—the Kramers–Kronig relations. The expression (3.11), which determines the values of the function ϵ for a purely imaginary argument $\omega=i\xi$ from the values of the function $\epsilon''(\omega)$ for real ω is a particular case of these relations; $\epsilon(i\xi)$ is a real quantity which decreases monotonically from ϵ_0 (the electrostatic dielectric constant) for $\xi=0$ to 1 for $\xi\to\infty$. It is these functions $\epsilon(i\xi)$ which appear in the expression (4.13). We can therefore state that the only macroscopic properties of bodies that determine the strength of the molecular interaction between them are the imaginary parts of their dielectric permeabilities‡.

Before we proceed with the discussion of the formula derived it is necessary to make the following general remark. If two bodies are separated by an empty gap, the electromagnetic forces which we have calculated are the only interaction forces between the bodies, but if the gap is occupied by some medium then we have the possibility of fluctuations in the medium which are due to non-electromagnetic oscillations (sound, for example) and these can contribute to the interaction. However, it will be shown in §5.2 that the contribution of these nonelectromagnetic forces is usually small.

---

† Most of the results given in §§ 4.2, 4.3 and 4.4 are taken from Lifshitz (1955).

‡ Equation (4.13) was derived on the assumption that all the media were isotropic and therefore its applicability to crystals depends on the possibility of neglecting the anisotropy of the dielectric permeability. Although this is admissible in the majority of cases, it should be remembered that in general the anisotropy of the bodies also leads to a specific effect—the appearance of a couple which tends to rotate the bodies with respect to one another.

If the bodies are identical ($\epsilon_1 = \epsilon_2$) then the expression under the integral is always positive† for all the terms of the sum in (4.13), and for given values of p and ξ_n it decreases monotonically as l increases. Hence it follows that $F > 0$ and $dF/dl < 0$, i.e. identical bodies attract one another irrespective of the size of the gap between them, while the attractive force decreases monotonically as the gap increases‡.

This statement is also valid for two different media which are separated by an empty gap ($\epsilon_3 = 1$). If the bodies differ and the space between them is filled by some fluid, then the force between them can be either attractive or repulsive (see below).

The general formula (4.13) is very complicated but it can be considerably simplified if we use the fact that the influence of temperature on the interaction force usually turns out to be negligible§.

Because of the presence of an exponential under the integral in (4.13) the important terms in the sum will be those for which $\xi_n \sim c/l$ or $n \sim ch/lkT$. For $lkT/ch \gg 1$ large values of n will thus be important and we can replace the sum by an integration over $dn = (h/2\pi kT)\,d\xi$. The temperature now drops out of the formula and we are left with the following result:

$$F = \frac{h}{2\pi^2 c^3} \int_0^\infty \int_1^\infty p^2 \xi^3 \epsilon_3^{3/2} \left\{ \left[\frac{(s_1+p)(s_2+p)}{(s_1-p)(s_2-p)} \exp\left(\frac{2p\xi}{c} l\sqrt{\epsilon_3} \right) - 1 \right]^{-1} + \left[\frac{(s_1+p\epsilon_1/\epsilon_3)(s_2+p\epsilon_2/\epsilon_3)}{(s_1-p\epsilon_1/\epsilon_3)(s_2-p\epsilon_2/\epsilon_3)} \exp\left(\frac{2p\xi}{c} l\sqrt{\epsilon_3} \right) - 1 \right]^{-1} \right\} dp\, d\xi. \quad (4.14)$$

This formula is valid for separations $l \ll ch/kT$; at room temperature this gives distances of up to $\sim 10^{-4}$ cm.

Equation (4.14) is still complex but it can be simplified further in two important limiting cases.

First we consider the limiting case of 'small' distances, by which we mean distances which are small compared to the wavelengths λ_0 which characterise the absorption spectra of the given bodies. The temperatures which are applicable for condensed bodies are certainly small compared to the $\hbar\omega$ which are important here (for example, in the visible region), and therefore the inequality $kTl/\hbar c \ll 1$ is known to be satisfied.

Because of the exponential factor $\exp(2p\xi l\sqrt{\epsilon_3}/c)$ in the denominators of the expression under the integral, those values of p for which $p\xi l/c \sim 1$ are dominant in the integration with respect to dp. In this case $p \gg 1$ and therefore we can put $s_1 \approx s_2 \approx p$ in the main terms. In this approximation the first term in curly brackets in (4.14) is zero, while the second term

† This is easily seen if we note that for $s = \sqrt{(\epsilon - 1 + p^2)}$ (where $p \geqslant 1$) the inequality $\epsilon p > s > p$ holds for $\epsilon > 1$, while for $\epsilon < 1$ $\epsilon p < s < p$.

‡ This assertion was made earlier by Hamaker (1937) on the basis of the assumption (which is in fact not valid) that the molecular forces should be additive.

§ When we talk of the influence of temperature, we exclude the temperature dependence related to the dependence of the dielectric permeability itself on temperature.

gives on the introduction of the variable of integration $x = 2lp\xi\epsilon_3^{1/2}/c$

$$F = \frac{\hbar}{16\pi^2 l^3}\int_0^\infty\int_0^\infty x^2\left[\frac{(\epsilon_1+\epsilon_3)(\epsilon_2+\epsilon_3)}{(\epsilon_1-\epsilon_3)(\epsilon_2-\epsilon_3)}e^x - 1\right]^{-1} dx\,d\xi \qquad (4.15)$$

(in this approximation the lower limit of integration with respect to dx is replaced by zero).

This force is proportional to the inverse cube of the distance, as would be predicted by the usual theory of van der Waals forces between two atoms. The functions $\epsilon(i\xi)-1$ decrease monotonically as ξ increases and tend to zero. Therefore values of ξ which lie beyond a certain ξ_0 make a negligible contribution to the integral; the condition that l should be small is that $l \ll c/\xi_0$.

To estimate the accuracy of the limiting behaviour derived above it is useful to have the next term of the expansion of the function $F(l)$. Calculation from the general formula (4.14) gives (for similar bodies separated by a vacuum, $\epsilon_1 = \epsilon_2 \equiv \epsilon$, $\epsilon_3 = 1$) the expression

$$-\frac{\hbar}{8\pi^2 c^2 l}\int_0^\infty \frac{\xi^2[\epsilon(i\xi)-1]^2}{\epsilon(i\xi)+1}d\xi, \qquad (4.16)$$

which must be added to (4.15). However, it is not possible to make a realistic estimate of the region of validity of the limiting law without knowing the form of the function $\epsilon(i\xi)$.

To an accuracy which is quite sufficient in practice eqn. (4.15) can be written in an even simpler form by neglecting unity compared to the term with e^x in the square brackets. [The accuracy of this reduction is connected to the fact that an integral of the form

$$\frac{a}{n!}\int_0^\infty \frac{x^n dx}{ae^x - 1} \qquad (4.17)$$

varies very little as a varies from ∞ to 1: from 1 to 1·2 for $n=2$, to 1·08 for $n=3$, to 1·04 for $n=4$, etc.] Then the integration with respect to dx can be carried out in an elementary way, and instead of (4.15) we get

$$F = \frac{\hbar\omega}{8\pi^2 l^3}, \bar{\omega} = \int_0^\infty \frac{(\epsilon_1(i\xi)-\epsilon_3(i\xi))(\epsilon_2(i\xi)-\epsilon_3(i\xi))}{(\epsilon_1(i\xi)+\epsilon_3(i\xi))(\epsilon_2(i\xi)+\epsilon_3(i\xi))}d\xi. \qquad (4.18)$$

The quantity $|\bar{\omega}|$ is some characteristic frequency for the absorption spectra of all three media.

We now turn to the opposite limiting case, that of 'large' distances; $l \gg \lambda\!\!\!^{-}_0$. We shall however suppose that the distances are not so large as to invalidate the inequality $lkT/\hbar c \ll 1$.

We introduce into the general formula (4.14) a new variable of integration $x = 2pl\xi/c$, but as the second variable we take not ξ (as above) but p:

$$F = \frac{\hbar c}{32\pi^2 l^4}\int_0^\infty\int_1^\infty \frac{x^3}{p^2}\epsilon_3^{3/2}\left\{\left[\frac{(s_1+p)(s_2+p)}{(s_1-p)(s_2-p)}\exp(x\sqrt{\epsilon_3}) - 1\right]^{-1}\right.$$
$$\left. + \left[\frac{(s_1+p\epsilon_1/\epsilon_3)(s_2+p\epsilon_2/\epsilon_3)}{(s_1-p\epsilon_1/\epsilon_3)(s_2-p\epsilon_2/\epsilon_3)}\exp(x\sqrt{\epsilon_3}) - 1\right]^{-1}\right\} dp\,dx$$
$$\epsilon = \epsilon(ixc/2pl), \quad s = \sqrt{(\epsilon - 1 + p^2)}.$$

Because of the presence of the factor $\exp(x\sqrt{\epsilon_3})$ in the denominators, values of $x \sim 1/\sqrt{\epsilon_3} \leqslant 1$ are the only ones important in the integration over dx, and since $p \geqslant 1$ the argument of the function ϵ for large l is nearly zero over the whole of the important range of the variables. Because of this we can replace ϵ_1, ϵ_2, ϵ_3 by their values for $\xi = 0$, i.e. the electrostatic dielectric constants. If we then replace $x\sqrt{\epsilon_{30}}$ by x, we finally obtain the following result

$$F = \frac{3\hbar c}{32\pi^2 l^4 \sqrt{\epsilon_{30}}} \int_0^\infty \int_1^\infty \frac{x^3}{p^2} \left\{ \left[\frac{(s_{10}+p)(s_{20}+p)}{(s_{10}-p)(s_{20}-p)} e^x - 1 \right]^{-1} + \frac{(s_{10}+p\epsilon_{10}/\epsilon_{30})(s_{20}+p\epsilon_{20}/\epsilon_{30})}{(s_{10}-p\epsilon_{10}/\epsilon_{30})(s_{20}-p\epsilon_{20}/\epsilon_{30})} e^x - 1 \right]^{-1} \right\} dp\, dx \quad . \quad (4.19)$$

$$s_{10} = \sqrt{(\epsilon_{10}/\epsilon_{30} - 1 + p^2)}, \quad s_{20} = \sqrt{(\epsilon_{20}/\epsilon_{30} - 1 + p^2)}$$

where ϵ_{10}, ϵ_{20}, ϵ_{30} are the electrostatic dielectric constant.

In accordance with the above-mentioned property of integrals of the type of (4.17) eqn. (4.19) can be written with considerable accuracy in the form:

$$F = \frac{3\hbar c}{16\pi^2 l^4 \sqrt{\epsilon_{30}}} \int_1^\infty \left\{ \frac{(s_{10}-p)(s_{20}-p)}{(s_{10}+p)(s_{20}+p)} + \frac{(s_{10}-p\epsilon_{10}/\epsilon_{30})(s_{20}-p\epsilon_{20}/\epsilon_{30})}{(s_{10}+p\epsilon_{10}/\epsilon_{30})(s_{20}+p\epsilon_{20}/\epsilon_{30})} \right\} \frac{dp}{p^2} . \quad . \quad . \quad (4.20)$$

Here only one integration remains and in principle this can be reduced to elementary functions; the result is however so unwieldy that in real calculations it is better to use numerical integration.

It was pointed out above that if the two bodies differ and the medium between them is filled with fluid the interaction can be either an attraction or a repulsion. Thus, from (4.18) it is clear that if $\epsilon_1 - \epsilon_3$ and $\epsilon_2 - \epsilon_3$ have opposite signs in the important frequency range, then $F < 0$, i.e. the bodies will repel each other for 'small' separations. At 'large' separations the forces are determined by the electrostatic values of the dielectric permeability; if $\epsilon_{10} - \epsilon_{30}$ and $\epsilon_{20} - \epsilon_{30}$ have the same sign $F > 0$, while if their signs differ $F < 0$. Furthermore, since the relative magnitudes of ϵ_{10}, ϵ_{20}, ϵ_{30} are not in general related to the behaviour of the functions $\epsilon_1(i\xi)$, $\epsilon_2(i\xi)$, $\epsilon_3(i\xi)$ in the frequency range which is important for these bodies, it is posssible in principle to have cases in which F changes sign at some value of l.

We now return to eqn. (4.19) and consider some of its special cases. In particular a simple result is obtained when both bodies are metals. For metals the function $\epsilon(i\xi)$ tends to infinity as $\xi \to 0$; therefore we must put $\epsilon_0 = \infty$. If we put $\epsilon_{10} = \epsilon_{20} = \infty$, we get

$$F = \frac{\hbar c}{16\pi^2 l^4 \sqrt{\epsilon_{30}}} \int_0^\infty \int_1^\infty \frac{x^3 dp dx}{p^2(e^x - 1)} = \frac{\pi^2}{240} \frac{\hbar c}{l^4 \sqrt{\epsilon_{30}}} . \quad . \quad (4.21)$$

This force is independent of the nature of the metal (a property that does not hold for small distances, where the magnitude of the interaction depends on the behaviour of the functions $\epsilon(i\xi)$ for all values of ξ and not just for $\xi = 0$). For $\epsilon_{30} = 1$ formula (4.21) coincides with that obtained by Casimir (1948) for this special case by considering the characteristic vibrations of the field in a gap between two walls which reflect ideally at all frequencies.

If both bodies are the same ($\epsilon_{10} = \epsilon_{20}$), (4.19) can be written in the form

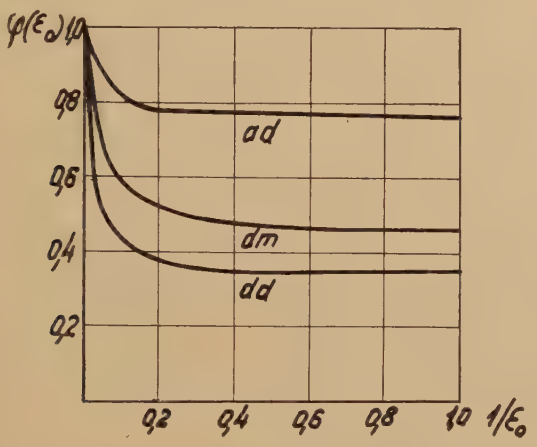
$$F = \frac{\pi^2}{240}\frac{\hbar c}{l^4}\frac{1}{\sqrt{\epsilon_{30}}}\left(\frac{\epsilon_{10}-\epsilon_{30}}{\epsilon_{10}+\epsilon_{30}}\right)^2 \phi_{dd}\left(\frac{\epsilon_{10}}{\epsilon_{30}}\right) \quad . \quad . \quad . \quad (4.22)$$

where $\phi_{dd}(x)$ is a function whose numerical values are given in fig. 10 (curve dd) for an argument ranging from 1 to ∞; in addition we note that $\phi_{dd}(0) = 0{\cdot}52$. For $x \to \infty$ ϕ_{dd} tends to 1 as

$$\phi_{dd}(x) \cong 1 - \frac{1{\cdot}11}{\sqrt{x}} \ln \frac{x}{7{\cdot}6} \; ;$$

for $x \to 1$ it tends to a finite limit 0·35 (according to the limiting law (4.35); see below).

Fig. 10



The same diagram shows the curve (dm) for the analogous function describing the attraction between the dielectric and a metal ($\epsilon_{20} = \infty$) by the following formula:

$$F = \frac{\pi^2}{240}\frac{\hbar c}{l^4}\frac{1}{\sqrt{\epsilon_{30}}}\frac{\epsilon_{10}-\epsilon_{30}}{\epsilon_{10}+\epsilon_{30}}\phi_{dm}\left(\frac{\epsilon_{10}}{\epsilon_{30}}\right). \quad . \quad . \quad . \quad (4.23)$$

For $\epsilon_3 \to \infty$ the expression (4.19) tends to zero. This means that when the gap is filled with liquid metal the interaction force at 'large' distances drops off as a higher power of $1/l$. This peculiar case is in principle of some interest although it is hardly of practical significance. To study it we must return to the initial expression (4.14) and in it take into account the real law according to which the dielectric permeability of the metal increases with a reduction in frequency.

For in a metal in the infra-red the function $\epsilon(\omega)$ is given with sufficient accuracy by the formula

$$\epsilon(\omega) = -4\pi e^2 N/m\omega^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.24)$$

where N is the density of the number of free electrons. When $\epsilon_3(i\xi) = 4\pi e^2 N/m\xi^2$ is substituted in (4.14) the exponential factors in the

denominators of the expression under the integral become

$$\exp\left\{\frac{2pl}{c}\sqrt{\left(\frac{4\pi e^2}{mN}\right)}\right\},$$

i.e. ξ drops out of the expression, while since p covers a range where $p \geqslant 1$ we conclude that the region of frequencies considered gives a contribution to the force F which decreases exponentially as the separation l increases.

In this case the main contribution to the interaction force is given by still smaller frequencies, where $\epsilon(\omega)$ is related to the usual electrical conductivity of the metal σ by the expression

$$\epsilon(\omega) = 4\pi i\sigma/\omega \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.25)$$

We substitute $\epsilon_3(i\xi) = 4\pi\sigma_3/\xi$ in the expression (4.14) (in the exponent and in the factor $\epsilon_3^{3/2}$; otherwise it is sufficient to put $\epsilon_3 = \infty$) and change the variable of integration from ξ to $x = 4pl\sqrt{(\pi\sigma_3\xi)}/c$, and we finally obtain

$$F = \frac{\hbar c^2}{2^7\pi^3\sigma_3 l^5}\int_0^\infty\int_1^\infty \frac{x^4}{p^2}\left\{\left[\left(\frac{\sqrt{(p^2-1)}+p}{\sqrt{(p^2-1)}-p}\right)^2 e^x - 1\right]^{-1} + \frac{1}{e^x-1}\right\} dp\,dx.$$

The calculations of the double integral (using the properties noted above of integrals of the type of (4.17)) gives the value 13·5, and as a result we have the following formula:

$$F = 0{\cdot}0034\hbar c^2/\sigma_3 l^5. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.26)$$

Thus, when two bodies are separated by a layer of liquid metal, the decrease in the molecular interaction as l^{-3} for 'small' distances becomes a decrease as l^{-5} for 'large' distances; though the onset of the latter is delayed by the presence of the small numerical coefficient.

It is natural to ask the real value of the wavelength λ_0 with which the distance l must be compared. However, the answer cannot be given in general as it depends on the real form of the spectral distribution of absorption by the given bodies (i.e. on the particular properties of the functions $\epsilon''(\omega)$). As an example we shall calculate the range of applicability of equation (4.21) for the interaction of two metals (which we shall take to be identical) separated by a vacuum.

Equation (4.21) is obtained from (4.14) by putting $\epsilon_1 = \epsilon_2 = \infty$ ($\epsilon_3 = 1$). If we wish to obtain the next term of the expansion as well, we must use the form (4.24) of the function $\epsilon(\omega)$, as this is valid over the frequency range which is important in the integration. (The region of still smaller frequencies, where $\epsilon(\omega)$ is given by (4.25), gives a very small contribution to the integral considered). After substituting $\epsilon(i\xi)$ in (4.14) we replace ξ by $xc/2pl$; an expansion of the expression under the integral in powers of $1/l$ gives

$$F = \frac{\hbar c}{32\pi^2 l^4}\left\{\frac{2\pi^4}{15} - \frac{c}{el}\sqrt{\left(\frac{m}{\pi N}\right)}\int_0^\infty \frac{x^4 e^x dx}{(e^x-1)^2}\int_1^\infty \frac{p^2+1}{p^4}\,dp\right\},$$

whence finally

$$F = \frac{\hbar c}{l^4}\frac{\pi^2}{240}\left\{1 - 7{\cdot}2\frac{c}{el}\sqrt{\left(\frac{m}{N}\right)}\right\}. \quad . \quad . \quad . \quad . \quad (4.27)$$

As a concrete example $N = 5{\cdot}9 \times 10^{22}$ for silver, and we find that the second term is small compared to the first provided $l \gg 0{\cdot}6\ 10^{-4}$ cm. We note that the next term of the expansion derived here could not have been obtained by the method which was used by Casimir to derive the first term.

The scope of this article does not include a review of the experimental data on van der Waals forces. We merely note that the first reliable measurements of the molecular attraction between solid bodies (on quartz)† were made by Deryagin and Abrikosova (1956) and Abrikosova (1957) and were in good agreement with theory. A detailed exposition and discussion of these data is given in the review articles by Deryagin *et al.* (1956, 1958). Similar measurements were also made by Kitchener and Prosser (1957) and de Jongh (1958).

4.3. *The Influence of Temperature*

All the formulae quoted in §4.2 are obtained on the assumption that $kTl/\hbar c \ll 1$, in accordance with which we retained only the first (zeroth) term of the expansion in powers of the temperature when we transformed from (4.13) to (4.14). To estimate the error which this introduced we need to find the next term of the expansion. We now do this for two identical metals separated by a vacuum.

The replacement of a sum by an integral in the derivation of (4.14) corresponds to the use of the first term of Euler's well known summation formula:

$$\sum_{n=0}^{\infty}{}' f(n) = \int_0^{\infty} f(n)\,dn + \frac{1}{12}f'(0) - \frac{1}{30{\cdot}4!}f'''(0) + \ldots \quad .$$

Here the role of the function $f(n)$ is played by the integral under the summation sign in (4.13). In the calculation we assume that l is small compared to $\hbar c/kT$, but still large compared to $(c/e)\sqrt{(m/N)}$, the characteristic quantity for the metal (see (4.27)). Then $f'(0) = 0$, $f'''(0) = 2$ and thus

$$F = \frac{\pi^2}{240}\frac{\hbar c}{l^4}\left[1 - \frac{48}{9}\left(\frac{lkT}{\hbar c}\right)^4\right]. \qquad . \quad . \quad . \quad . \quad (4.28)$$

† The case of quartz has certain peculiarities because of the specific properties of its absorption spectrum. Quartz absorbs strongly in the ultra-violet (from approximately $0{\cdot}15\ \mu$) and in the infra-red (from several μ) and between these wavelengths it is transparent. For separations l which lie in the transparent region a reasonable estimate of the force F may be obtained by taking l to be small compared to λ at the right-hand boundary and large compared to λ at the left-hand boundary of the region. The contribution of the ultra-violet absorption band to F can be estimated from (4.22) by putting $\epsilon_{10} = \epsilon_{20} = \epsilon_0$ ($\epsilon_{30} = 1$) equal to the square of the refractive index in the transparent optical region. The contribution of the infra-red region is given by (4.18); in order of magnitude it is less by a factor $l\omega_0/c$ (ω_0 is the infra-red absorption frequency). Thus to estimate the attractive force we can use (4.22) with the optical (instead of the electrostatic) value of the dielectric permeabilities substituted for ϵ_0. This estimate is low for larger separations and high for smaller ones.

Thus at room temperature the correction term is small for $l \lesssim 5 \times 10^{-4}$ cm; comparison with the criterion derived from (4.27) shows that there exists a region of validity of the formula (4.21).

When $lkT/\hbar c \gg 1$ we need to keep only the first of the terms of the sum (4.13). However, we cannot immediately put $n=0$ because of the indeterminacy which this introduces (the factor $\xi_n{}^3$ tends to zero, but the intergal over dp diverges). This difficulty can be avoided by first replacing p by a new variable of integration $x = 2p\xi_n l\sqrt{\epsilon_{30}}/c$ (as a result of which the factor $\xi_n{}^3$ disappears). If we then put $\xi_n = 0$ we find

$$F = \frac{kT}{16\pi l^3}\int_0^\infty x^2 \left[\frac{(\epsilon_{10}+\epsilon_{30})(\epsilon_{20}+\epsilon_{30})}{(\epsilon_{10}-\epsilon_{30})(\epsilon_{20}-\epsilon_{30})}e^x - 1\right]^{-1} dx$$
$$\cong \frac{kT}{8\pi l^3}\frac{(\epsilon_{10}-\epsilon_{30})(\epsilon_{20}-\epsilon_{30})}{(\epsilon_{10}+\epsilon_{30})(\epsilon_{20}+\epsilon_{30})}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.29)$$

Thus at sufficiently large distances the decrease in the interaction slows down and again goes as l^{-3} with a coefficient which now depends both on the temperature and on the electrostatic value of the dielectric permeabilities.

All the other terms of the sum (4.13) decrease exponentially for large $lkT/\hbar c$. Thus, for two metals separated by a vacuum, the corrected force is

$$F = \frac{kT}{8\pi l^3}\left[1 + 2\left(\frac{4\pi kTl}{\hbar c}\right)^2 \exp\left(-\frac{4\pi kTl}{\hbar c}\right)\right]. \quad . \quad . \quad (4.30)$$

4.4. *Interaction of Individual Atoms*

We now show how it is possible to go from the macroscopic formula (4.14) to the interaction of individual atoms in a vacuum. To do this we formally suppose both bodies to be sufficiently rarefied. From the point of view of macroscopic electrodynamics this means that their dielectric permeabilities are close to 1, i.e. the differences $\epsilon_1 - 1$ and $\epsilon_2 - 1$ are small.

We begin with 'small distances'. From (4.15) with $\epsilon_3 = 1$ we have with the necessary accuracy:

$$F = \frac{\hbar}{64\pi^2 l^3}\int_0^\infty\int_0^\infty x^2 e^{-x}(\epsilon_1 - 1)(\epsilon_2 - 1)\,dx\,d\xi$$
$$= \frac{\hbar}{32\pi^2 l^3}\int_0^\infty [\epsilon_1(i\xi) - 1][\epsilon_2(i\xi) - 1]\,d\xi \quad . \quad . \quad (4.31)$$

Expressing $\epsilon(i\xi)$ in terms of $\epsilon''(\omega)$ along the real ω-axis according to (3.11) we have

$$\int_0^\infty [\epsilon_1(i\xi) - 1][\epsilon_2(i\xi) - 1]\,d\xi = \frac{4}{\pi^2}\int_0^\infty\int_0^\infty\int_0^\infty \frac{\omega_1\omega_2\epsilon_1''(\omega_1)\epsilon_2''(\omega_2)}{(\omega_1{}^2+\xi^2)(\omega_2{}^2+\xi^2)}\,d\xi\,d\omega_1\,d\omega_2$$
$$= \frac{2}{\pi}\int_0^\infty\int_0^\infty \frac{\epsilon_1''(\omega_1)\epsilon_2''(\omega_2)}{\omega_1+\omega_2}\,d\omega_1\,d\omega_2$$

whence for the force F we find

$$F = \frac{\hbar}{16\pi^3 l^3}\int_0^\infty\int_0^\infty \frac{\epsilon_1''(\omega_1)\epsilon_2''(\omega_2)}{\omega_1+\omega_2}\,d\omega_1\,d\omega_2. \quad . \quad . \quad (4.32)$$

This force corresponds to an atomic interaction with an energy†

$$U = -\frac{3\hbar}{8\pi^4 R^6 N_1 N_2}\int_0^\infty\int_0^\infty \frac{\epsilon_1''(\omega_1)\epsilon_2''(\omega_2)}{\omega_1+\omega_2}\,d\omega_1\,d\omega_2 \quad . \quad (4.33)$$

where R is the distance between the atoms, N_1, N_2 are the numbers of atoms per unit volume in the two bodies. The imaginary part of the dielectic permeability is related to the spectral density of the 'oscillator strengths' $f(\omega)$ which are known from spectroscopy by the equation

$$\omega\epsilon''(\omega) = \frac{2\pi^2 e^2}{m} Nf(\omega)$$

(see, for example, Landau and Lifshitz, 1960, §62). Substituting this in (4.33) we find

$$U(R) = -\frac{3\hbar e^4}{2m^2R^6}\int_0^\infty\int_0^\infty \frac{f_1(\omega_1)f_2(\omega_2)}{\omega_1\omega_2(\omega_1+\omega_2)}\,d\omega_1\,d\omega_2. \quad . \quad (4.34)$$

This expression is identical with the well-known London formula (1930) which was obtained by ordinary perturbation theory applied to the dipole interaction of two atoms. For example, suppose we are considering the interaction of two hydrogen atoms. We use the expression

$$f_{0n} = \frac{2m}{\hbar^2}(E_n - E_0)|x_{0n}|^2$$

for the oscillator strength for the transition between the states E_n and E_0 (x_{0n} is the corresponding matrix element of the coordinates of the electron in the atom), and when we transform from an integration over frequencies to a summation over the energy levels of the atom we get the London formula for hydrogen atoms:

$$U(R) = -\frac{6e^4}{R^6}\sum_{n,m}\frac{|x_{0n}|^2|x_{0m}|^2}{E_n - E_0 + E_m - E_0}.$$

Thus we see how this 'microscopic' formula is reproduced from a purely macroscopic theory.

For 'large' distances the formula for the attractive force between two rarefied bodies takes the form

$$F = \frac{\hbar c}{32\pi^2 l^4}(\epsilon_{10}-1)(\epsilon_{20}-1)\int_0^\infty x^3 e^{-x}\,dx\int_1^\infty \frac{1-2p^2+2p^4}{8p^6}\,dp$$

or

$$F = \frac{\hbar c}{l^4}\frac{23}{640\pi^2}(\epsilon_{10}-1)(\epsilon_{20}-1). \quad . \quad . \quad . \quad . \quad (4.35)$$

† If the potential energy of the interaction between molecules 1 and 2 is $U = -aR^{-6}$, the total energy of the interactions in pairs of all the molecules in the two half-spaces separated by a gap l is

$$\bar{U} = -a\pi N_1 N_2/12l^2.$$

The force F is

$$F = -\frac{d\bar{U}}{bl} = -\frac{a\pi N_1 N_2}{6l^3}.$$

This explains the correspondence of eqns. (4.32) and (4.33)

This force corresponds to the interaction of two atoms with an energy

$$U = -\frac{23\hbar c}{4\pi R^7}\alpha_1\alpha_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.36)$$

where α_1, α_2 are the static polarizabilities of the atoms ($\epsilon_0 = 1 + 4\pi N\alpha$). Formula (4.36) coincides with the result of the quantum mechnical calculation by Casimir and Polder (1948) of the attraction between two atoms when the distance between them is sufficiently large for retardation effects to be important.

Similarly, by considering just one of the bodies, say body 2, as a rarefied medium, we can find the interaction between individual molecules and a condensed body. Thus, for a molecule which is at a 'large' distance l from the surface of the body we find the following formula for the interaction energy

$$U(l) = \frac{3\hbar c\alpha_2}{8\pi l^4}\frac{\epsilon_{10}-1}{\epsilon_{10}+1}\phi_{ad}(\epsilon_{10}) \quad . \quad . \quad . \quad . \quad . \quad (4.37)$$

where

$$\phi_{ad}(\epsilon) = \frac{\epsilon+1}{\epsilon-1}\left\{\frac{1}{3} + \epsilon + \frac{4-(\epsilon+1)\sqrt{\epsilon}}{2(\epsilon-1)}\right.$$
$$\left. - \frac{\operatorname{Arsh}\sqrt{(\epsilon-1)}}{2(\epsilon-1)^{3/2}}[1+\epsilon+2\epsilon(\epsilon-1)^2] + \frac{\epsilon^2}{\sqrt{(\epsilon+1)}}\left(\operatorname{Arsh}\sqrt{\epsilon} - \operatorname{Arsh}\frac{1}{\sqrt{\epsilon}}\right)\right\}. \quad . \quad . \quad . \quad (4.38)$$

The function ϕ_{ad} is given graphically in fig. 10. For $\epsilon \to \infty$ it tends to 1, while for $\epsilon \to 1$ it tends to $23/30 = 0{\cdot}77$. The expression for $\epsilon_{10} \to \infty$

$$U(l) = \frac{3\alpha_2\hbar c}{8\pi l^4} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.39)$$

coincides with the results of Casimir and Polder (1948) for the interaction energy of an atom with a metal wall.

We now consider the interaction of two atoms in a liquid (Pitaevskii 1959). We suppose that both bodies are weak solutions in the same solvent of atoms of different kinds with concentrations (numbers of particles per cm^3) N_1 and N_2. We also suppose that the gap is filled with pure solvent. For small concentrations of dissolved atoms the dielectric permeabilities ϵ_1, ϵ_2 of the solutions differ little from that of the pure solvent, which we denote by $\epsilon_3 = \epsilon$. To first order in the concentration we have

$$\epsilon_1 = \epsilon + N_1\left(\frac{\partial\epsilon_1}{\partial N_1}\right)_{N_1=0}, \quad \epsilon_2 = \epsilon + N_2\left(\frac{\partial\epsilon_2}{\partial N_2}\right)_{N_2=0}.$$

From the formula (4.15) for the force at 'small' distances we retain only those terms of the same order of smallness, and (analogously to the transform to (4.31)) we find:

$$F(l) = \frac{\hbar}{32\pi^2 l^3}N_1N_2\int_0^\infty \left(\frac{\partial\epsilon_1(i\xi)}{\partial N_1}\right)_{N_1=0}\left(\frac{\partial\epsilon_2(i\xi)}{\partial N_2}\right)_{N_2=0}\frac{d\xi}{\epsilon^2(i\xi)}.$$

This force corresponds to an interaction energy of the dissolved atoms equal to

$$U(R) = -\frac{3\hbar}{16\pi^3 R^6}\int_0^\infty \left(\frac{\partial\epsilon_1(i\xi)}{\partial N_1}\right)_{N_1=0}\left(\frac{\partial\epsilon_2(i\xi)}{\partial N_2}\right)_{N_2=0}\frac{d\xi}{\epsilon^2(i\xi)}. \quad (4.40)$$

Analogously the energy for 'large' distances is found to be

$$U(R) = -\frac{23\hbar c}{64\pi^3\epsilon_0^{3/2}R^7}\left(\frac{\partial\epsilon_{10}}{\partial N_1}\right)_{N_1=0}\left(\frac{\partial\epsilon_{20}}{\partial N_2}\right)_{N_2=0}. \quad (4.41)$$

We see that when the dissolved molecules interact strongly with the solvent the interaction forces between them are no longer determined by their polarisability.

The interaction of small spherical particles in a liquid is another interesting example. Let both bodies be such an emulsion, made up of spherical particles of volume v with dielectric permeability ϵ' in a liquid with dielectric permeability ϵ. As before the gap is filled with pure solvent. For $Nv \ll 1$ (N is the number per unit volume) the dielectric permeability of the emulsion is of the form

$$\epsilon_1 = \epsilon_2 = \epsilon + 3Nv\frac{(\epsilon'-\epsilon)\epsilon}{\epsilon'+2\epsilon}$$

(see, for example, Landau and Lifshitż, 1960, §9). By using the fact that $\epsilon_1 - \epsilon$ and $\epsilon_2 - \epsilon$ are both small we find by a similar method to that used above the following expressions for the interaction energy of the particles of the emulsion:

$$U(R) = -\frac{27\hbar v^2}{16\pi^3 R^6}\int_0^\infty \left[\frac{\epsilon'(i\xi)-\epsilon(i\xi)}{\epsilon'(i\xi)+2\epsilon(i\xi)}\right]^2 d\xi, \quad R \ll \lambda_0, \quad (4.42)$$

$$U(R) = -\frac{207v^2\hbar c}{64\pi^3\sqrt{\epsilon_0}R^7}\left(\frac{\epsilon_0'-\epsilon_0}{\epsilon_0'+2\epsilon_0}\right)^2, \quad R \gg \lambda_0 \quad (4.43)$$

The dimensions of the particles themselves must be small compared to their separation R (but not necessarily compared to λ_0).

§ 5. A Thin Film on the Surface of a Solid Body

5.1. *The Chemical Potential of the Film*

The general theory of van der Waals forces developed above can also be used for the calculation of the thermodynamic quantities of a thin liquid film which lies on the surface of a solid body; the thickness l of the film is of course supposed to be large compared to the interatomic distances.

Equation (3.25), which was derived above, expresses the chemical potential per unit mass of a liquid in terms of the Green functions of the relevant fluctuations of the electromagnetic field. However, this formula is inconvenient for two reasons: first, it contains over the whole frequency range the quantity $\partial\epsilon/\partial\rho$ which has not been studied at all experimentally; and second, it gives the chemical potential ζ as a function of the density ρ, while it is usually necessary to know ζ as a function of the pressure p.

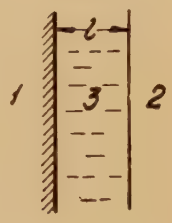
We consider a film 3 lying on the surface of a solid body 1 and in equilibrium with its vapour 2 (fig. 11). As far as its electromagnetic properties are concerned, we shall treat the vapour as a vacuum, i.e. we shall always take its dielectic permeability ϵ_2 to be equal to 1.

According to the condition of mechanical equilibrium, the normal component σ_{xx} of the stress tensor must be continuous at the surface of the film. Hence we find the equation

$$p = p_0(\rho, T) - \bar{\sigma}_{xx}$$

where p is the vapour pressure, $p_0(\rho, T)$ is the pressure of the bulk at the given density and temperature, and $\bar{\sigma}_{xx}$ denotes the sum of all the terms except the first in the expression (3.24) for the stress tensor in the film.

Fig. 11



By solving this equation for ρ we find the density in the form†

$$\rho = \rho_0(p + \bar{\sigma}_{xx}, T).$$

Substitution of this expression in eqn. (3.25) for the chemical potential gives

$$\zeta = \zeta_0(p + \bar{\sigma}_{xx}, T) + \frac{T}{4\pi} \sum_{n=0}^{\infty}{}' \frac{\partial \epsilon}{\partial \rho} \mathbb{D}_{ii}^{E}(\mathbf{r}, \mathbf{r}; \xi_n)$$

where now $\zeta_0(\rho, T)$ is the chemical potential of the bulk liquid. We expand ζ_0 in powers of the small quantity $\bar{\sigma}_{xx}$, and using the thermodynamic relation $(\partial\zeta/\partial p)_T = 1/\rho$ we find

$$\zeta(p, T) = \zeta_0(p, T) + \frac{1}{\rho} \bar{\sigma}_{xx} + \frac{T}{4\pi} \sum_{n=0}^{\infty}{}' \frac{\partial \epsilon}{\partial \rho} \mathbb{D}_{ii}^{E}.$$

Finally, by substituting here the expression for $\bar{\sigma}_{xx}$ from (3.24), we find that the term in $\partial\epsilon/\partial\rho$ drops out and we are left with

$$\zeta(p, T) = \zeta_0(p, T) + (1/\rho)\sigma_{xx}'$$

where σ_{xx}' is a component of the 'contracted' stress tensor (4.2). This quantity is constant across the film (because the momentum flux is constant), and the force $F(l)$, by (4.3), is directly determined by it.

We introduce the notation μ for the 'van der Waals part' of the chemical potential of the film per unit volume of the liquid:

$$\zeta = \zeta_0 + \mu/\rho. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.1)$$

† $\bar{\sigma}_{xx}$ is also a function of ρ, but since it is a small correction to the pressure we shall put $\rho = \rho_0(p, T)$.

From the above argument we have

$$\mu = \sigma_{xx}' = F(l).$$

As l tends to ∞, i.e. for the bulk liquid, μ tends to zero.

Thus we do not need to carry out any new calculations in order to determine the quantity μ. It is determined by the formulae obtained above for $F(l)$ (the general formula (4.13)) and the limiting cases considered in §§ 4.2, 4.3) simply by putting $\epsilon_2 = 1$.

The function $\mu(T, l)$ determines all the thermodynamic quantities of the film. So, if the film is in equilibrium with vapour at pressure p, the condition that the chemical potentials of the liquid and the vapour should be equal gives the well known equation

$$\mu(l) = \frac{kT\rho}{m} \ln \frac{p}{p_{\text{sat}}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.2)$$

where m is the mass of a molecule, and p_{sat} is the saturated vapour pressure above the surface of the bulk liquid at the given temperature T. The equilibrium thickness of the film is determined by this equation.†

If we consider a liquid film under gravity on a vertical wall, then $p = p_{\text{sat}} \exp(-mgz/kT)$ (where z is the height above the level of the liquid in the vessel) and from (5.2) we have the equation

$$\mu(l) + \rho g z = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.3)$$

which determines the profile of the film, i.e. its thickness as a function of height.

For 'small' (in the sense of § 4.2) film thicknesses we have the limiting behaviour (cf. 4.18)

$$\mu = \frac{\hbar\bar{\omega}}{8\pi^2 l^3}, \quad \bar{\omega} = \int_0^\infty \frac{(\epsilon_3 - 1)(\epsilon_3 - \epsilon_1)}{(\epsilon_3 + 1)(\epsilon_3 + \epsilon_1)} d\xi. \quad . \quad . \quad . \quad (5.4)$$

For 'large' thicknesses $\mu(l)$ is proportional to l^{-4} with a coefficient determined (in accordance with (4.20)) by the electrostatic dielectric constants of the film (ϵ_{30}) and of the solid body (ϵ_{10}); the sign of μ is the same as that of $\epsilon_{30} - \epsilon_{10}$‡. The function $\mu(l)$ may change sign and be non-monotonic (cf. the similar note on the force $F(l)$ on page 192). Non-monotonic behaviour of $\mu(l)$ in a certain range of values of l is in general related to a change in sign of the differences $\epsilon_3(i\xi) - \epsilon_1(i\xi)$ in the wavelength range $\lambda \sim l$.

As well as the potential μ it is convenient to use the 'effective surface tension coefficients' α at the boundary of the solid body and the vapour 2

---

† In the deduction of (5.2) the formula

$$\zeta_{\text{vap}} = \zeta_0(p_{\text{sat}}, T) + \frac{kT}{m} \ln \frac{p}{p_{\text{sat}}}$$

for the chemical potential is used, and the liquid is taken to be incompressible, i.e. we neglect the dependence of the chemical potential of the bulk liquid on pressure.

‡ Provided the important dispersion of the dielectric constant is not already apparent (as it is for water) at very long wavelengths.

to describe the properties of the film. This coefficient takes into account the existence of a liquid layer betewen them. This can be done formally by using the absorption theory relation

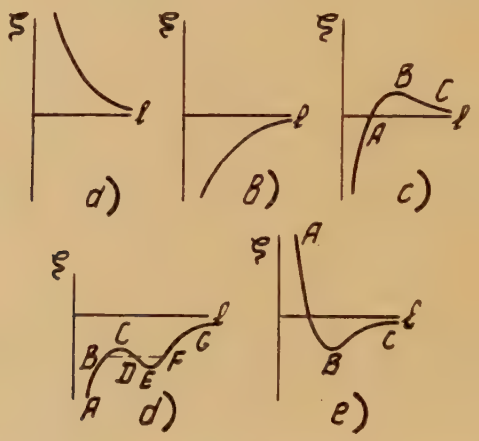
$$\gamma = -\left(\frac{\partial \alpha}{\partial \zeta'}\right)_T,$$

where γ is the surface concentration of adsorbed particles (the number per cm^2), ζ' is the chemical potential per particle (see, for example, Landau and Lifshitz (1958), §144). For the definition of μ taken here (the liquid is considered to be incompressible) this expression is written in the form

$$l = -\left(\frac{\partial \alpha}{\partial \mu}\right)_T, \qquad \ldots\ldots\ldots \quad (5.5)$$

which is applicable for both macroscopically thick 'wetting' films and adsorption films of 'molecular thickness': in the latter case, of course, l has only an indirect interpretation as a quantity proportional to the

Fig. 12



surface concentration ($l = m\gamma/\rho$, m is the mass of a molecule). Integrating (5.5) and using the fact that as $l \to \infty$ the function $\alpha(l)$ must tend to $\alpha_{13} + \alpha_{32}$, the sum of the surface tensions at the boundaries of the phases 1, 3 and 3, 2, we have

$$\alpha(l) = \int_l^\infty l \frac{d\mu}{dl}\, dl + \alpha_{13} + \alpha_{32}. \qquad \ldots\ldots \quad (5.6)$$

We also note that the necessary condition for the thermodynamic stability of the film is that the inequality

$$\left(\frac{\partial \mu}{\partial l}\right)_T > 0 \qquad \ldots\ldots\ldots \quad (5.7)$$

should be satisfied.

If (5.2) is satisfied by several values of l, the stable state of the film corresponds to that for which α is a minimum; higher values of α then correspond to metastable states.

We consider several typical cases which can arise for various functions $\mu(l)$.

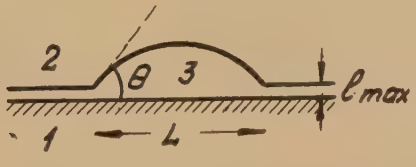
(*a*) If $\mu(l)$ is a monotonically decreasing function which is always positive (fig. 12 (*a*)), the liquid will not wet a solid surface and no film will be formed. We emphasize that we are now considering only macroscopically thick films, to which all the theory developed here refers. As far as adsorption in the narrow sense of the word is concerned, this, as is known, always takes place to a certain extent. This corresponds to the fact that, whatever the variation of the function $\mu(l)$ in the range of molecular dimensions (not shown in fig. 12), it must eventually tend to $-\infty$ for $l \to 0$ as $\mu \sim \ln l$ which corresponds to a 'weak solution' of adsorbed molecules on the surface.

(*b*) If $\mu(l)$ increases monotonically and is everywhere negative (fig. 12 (*b*)), this corresponds to a liquid which completely wets a solid surface and forms (depending on the vapour pressure above it) a stable film of any thickness. In particular, the thickness of a film formed on a vertical wall tends to zero as $z \to \infty$; the decrease initially goes as $l \sim z^{-1/4}$, and then as $l \sim z^{-1/3}$.

However, in this case too the liquid can be non-wetting if the behaviour of $\mu(l)$ in the microscopic range leads to smaller values of the surface tension α†; then molecular adsorption and not a wetting film will be stable.

(*c*) $\mu(l)$ passes through zero and has a maximum as shown in fig. 12 (*c*). With the same proviso as in (*b*) wetting will occur, but with the formation of a film which is stable only for thicknesses less than a certain limit. In equilibrium with the saturated vapour there will be a film of finite thickness, corresponding to the point A. This state is separated from the other stable state—the equilibrium of the solid body with the bulk liquid—by a metastable region AB and an unstable region BC.

Fig. 13



A curve $\mu(l)$ of this type must lead to interesting features in the formation of an angle of contact θ between the liquid drop and the solid body. In this case the drop is in equilibrium with a film of a finite thickness $l_{\max}$ (fig. 13 and by the usual elementary formula we have

$$\cos\theta = \frac{\alpha(l_{\max}) - \alpha_{13}}{\alpha_{23}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.8)$$

where $\alpha(l_{\max})$ (with $\alpha(l)$ from (5.6)) plays the role of the surface tension

† This could be a high peak in $\mu(l)$ in the molecular range of 'thicknesses'.

between phases 1 and 2. Since the first term in (5.6) is small, (5.8) gives

$$\theta^2 \cong -\frac{2}{\alpha_{23}}\int_{l_{\max}}^{\infty} l\frac{d\mu}{dl}\,dl = \frac{2}{\alpha_{23}}\int_{l_{\max}}^{\infty}\mu dl \quad . \quad . \quad . \quad . \quad (5.9)$$

By interpolating between $\mu \sim l^{-3}$ and $\mu \sim l^{-4}$ we can get an estimate for

$$\theta \sim \frac{1}{10 l_{\max}}\sqrt{\left(\frac{\hbar|\omega|}{\alpha_{23}}\right)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.10)$$

with $\bar{\omega}$ taken from (5.4). Thus, for $\hbar\bar{\omega} \sim 10\,\text{ev}$, $\alpha_{23} \sim 20\,\text{erg/cm}^2$, $l_{\max} \sim 5 \times 10^{-5}\,\text{cm}$, we get $\theta \sim 0{\cdot}1°$.

Thus in this case the angle of contact must be finite but it is very small (as distinct from the value $\theta = 0$ for complete wetting and $\theta \sim 1$ for the usual cases of non-wetting). It is understood that this type of behaviour can be observed only when the thickness of the drop is large compared to that of the film, i.e. we must have $L\theta \gg l_{\max}$, where L is the size of the drop (fig. 13).

(*d*) A curve of the type shown in fig. 12 (*d*)) corresponds to a film which is unstable over a certain range of thicknesses. The straight line BF, which cuts off equal areas BCD and DEF, joins the points B and F which have (for equal μ) equal values of α (as can easily be seen from (5.6)). The branches AB and FG correspond to stable films; the interval CE is unstable, and the intervals BC and EF are metastable.

In this case both boundaries of the range of instability (points B and F) correspond to macroscopic film thicknesses. Instability over the range from some macroscopic thickness to a molecular layer would correspond to the curve shown in fig. 12 (*e*) (for $l \to 0$ this, as does that in fig. 12 (*a*), tends to $-\infty$). Actually, however, such a curve would be more likely to lead simply to non-wetting. The boundary of stability would correspond to a point on the branch BC such that a horizontal line through it would cut off equal areas under the upper and above the lower parts of the curve However, the latter area related to the van der Waals forces, would be small compared to the former, because of the considerably greater forces at molecular distances. This means that the surface tension over the whole of the branch BC will be greater than that which corresponds to molecular adsorption on the surface of the solid body, and therefore the film will be metastable.

5.2. *Non-electromagnetic Forces*

As has already been noted at the beginning of § 4.2, forces of non-electromagnetic origin, as well as van der Waals forces, make a certain contribution to the chemical potential of the film, but this contribution is usually small. We now quote the corresponding estimates without going through the calculations in detail.

At absolute zero acoustic fluctuations (in an acoustically non-dispersive medium) give a contribution to the chemical potential

$$\mu_{ac} \sim \hbar u/l^4$$

where u is the velocity of sound†. This is to be compared to the electromagnetic part $\mu_{em} \sim \hbar c/l^4$ for $l \gg \lambda_0$ or $\mu_{em} \sim \hbar c/l^3\lambda_0$ for $l \ll \lambda_0$. It is clear that $\mu_{ac} \gg \mu_{em}$ for all distances large compared to atomic dimensions, which is also a condition that the theory developed here should be valid.

For non-zero temperatures the other limiting case for μ_{ac} usually holds, i.e. the influence of temperature is predominant. The corresponding criterion is the value of the ratio $lkT/\hbar u$. The condition $lkT/\hbar u \gg 1$ (as is the condition $lkT/\hbar c \gg 1$ in the electromagnetic case) is essentially the condition for classical behaviour ($\hbar\omega \ll kT$ with $\omega \sim l/u$ or $\omega \sim l/c$). It is therefore clear that the contribution to μ need not contain $\hbar$, and hence from dimensional considerations it is clear that

$$\mu_{ac} \sim kT/l^3$$

(cf. eqn. (3.19)). This is comparable with μ_{em} only for $l \sim \hbar c/kT$ which is a distance so large that μ itself has become very small.

The same applies to the contribution from the surface vibrations. The dependence of frequency on wave vector κ for capillary waves on the surface of a liquid layer of depth l is given by the well-known formula

$$\omega^2 = \frac{\alpha\kappa^3}{\rho} \operatorname{th} \kappa l$$

where α is the surface tension (see, for example, Landau and Lifshitz (1959), § 61); for a deep layer of liquid ($l \to \infty$) $\omega^2 = \alpha\kappa^3/\rho$. Calculating the energy of the zero point vibrations (with the subtraction of this energy as $l \to \infty$), we find that at absolute zero the corresponding contribution to the chemical potential is

$$\mu_{sur} \sim \frac{\hbar}{l^{9/2}} \sqrt{\frac{\alpha}{\rho}}.$$

In fact, however, the other limiting case is valid, i.e. the classical condition $(\hbar/kT)\sqrt{(\alpha/\rho)}l^{-3/2} \ll 1$ applies; on general statistical grounds calculation gives a contribution of the same order of magnitude $\mu_{sur} \sim kT/l^3$ as in the acoustic case‡.

To explain the properties of helium films other mechanisms related to inhomogeneities in the distribution of liquid density across the film have been proposed by several authors. In its crudest form the corresponding

† This expression is analogous to $\mu \sim \hbar c/l^4$ obtained for the electromagnetic contribution (in a non-dispersive medium). This can be derived by, for example, summing the energy of the zero point acoustic vibrations in the gap (of width l) similarly to the method used by Casimir (1948) for the zero point electromagnetic vibrations. We note that Atkins' (1954) result, which predicted that μ_{ac} should depend on the film thickness as l^{-2}, was due to an incorrect cut-off of a divergent integral.

‡In all this we are making literal estimates, but it should be remembered that really the expressions for μ_{sur} and μ_{ac} contain (as is shown by a more detailed analysis) numerical coefficients which are as small as those which appear in the electromagnetic part μ_{em}. The appearance of comparatively small numerical coefficients is a general characteristic of the theory developed here.

calculation treated the helium in the film as an ideal gas whose particle wave functions have nodes at the wall and at the surface of the film. This model leads to a sharply inhomogeneous density distribution with a maximum in the centre, and to a contribution to the chemical potential μ which is proportional to l^{-2}. This treatment is however quite inadmissible (as was pointed out by Mott (1949)), since the interaction between the atoms smooths out the wave function of the ground state of the system and the density inhomogeneities extend (in the body of the liquid) only over distances of the order of the atomic separation. The contribution to the chemical potential from this inhomogeneity decreases exponentially as the film thickness is increased.

The same dependence on film thickness applies to the contribution from the specific properties of helium below the λ-point (superfluidity). It is only in the immediate neighbourhood of the λ-point, where the density of the superfluid component is very small, that the inhomogeneity in the distribution of the latter produces any noticeable effects (see Ginsburg and Pitaevskii 1958). Even at 0·01° from the λ-point the decrement of the exponential decay becomes comparable to the interatomic distance. The result of Franchetti (1957), who derived a contribution to the chemical potential which was proportional to l^{-2}, is due to the inadequacy of the model of non-interacting elementary excitations in helium which he used.

5.3. *Films of Liquid Helium*

We consider in particular liquid helium films, on which there is already an extensive literature.

For helium films the general formula (4.14) can be considerably simplified by noting that the dielectric permeability of liquid helium is very close to unity, i.e. the difference $\epsilon_3(i\xi)-1$ is small. Carrying out the corresponding expansions in (4.14) we get

$$\mu(l)=-\frac{\hbar}{8\pi^2c^3}\int_0^\infty\int_1^\infty\left\{\frac{s_1-p}{s_1+p}+(2p^2-1)\frac{p\epsilon_1-s_1}{p\epsilon_1+s_1}\right\}(\epsilon_3-1)\xi^3 \exp(-2p\xi l/c)\,dp\,d\xi,$$

$$s_1=\sqrt{[\epsilon_1(i\xi)-1+p^2]}. \qquad (5.11)$$

However, calculation even from this simplified formula is hindered by the need to know the form of the function $\epsilon(i\xi)$ for the liquid helium and for the solid wall over a wide frequency range, in particular in the extreme ultraviolet: in the integral (5.11) the important range of wavelengths are those for which $\lambda\sim l$, while in fact the thickness of helium films is of the order of 10^{-6} cm.

For further simplification of (5.11) it is a reasonable approximation to use the fact that the main absorption band of helium lies in the extreme ultraviolet, while those for the solid body of the wall (metals, quartz) are at much lower frequencies. In other words we shall assume that the function $\epsilon_3(i\xi)$ practically coincides with the electrostatic value ϵ_{30} over the whole range of variation of ξ in which $\epsilon_1(i\xi)-1$ (and with it the whole of the

expression under the integral in (5.11)) is not too small. Then $\epsilon_3 - 1$ can be taken in front of the integral sign, and the remaining integral is transformed as in the limiting case of small thickness (l is small compared to the wavelengths λ_0 in the main absorption band of the solid body). We replace p by the variable of integration $x = 2p\xi l/c$ and use the fact that values of $x \sim 1$ correspond to large values of p to replace the curly brackets in (5.11) by $2p^2(\epsilon_1 - 1)/(\epsilon_1 + 1)$. As a result we find

$$\mu(l) = -\frac{\hbar\bar{\omega}(\epsilon_{30} - 1)}{16\pi^2 l^3} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.12)$$

where we have used the notation

$$\bar{\omega} = \int_0^\infty \frac{\epsilon_1(i\xi) - 1}{\epsilon_1(i\xi) + 1}\, d\xi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.13)$$

i.e. an average frequency which is characteristic of the given solid body.

We note that the function $[\epsilon(\omega) - 1]/[\epsilon(\omega) + 1]$ has the same analytic properties in the upper half-plane of the complex variable ω as the function $\epsilon(\omega) - 1$. This is sufficient to enable us to use the same formula to transform the integral along the imaginary axis to one along the real axis as was valid for the function $\epsilon(\omega) - 1$ (see Landau and Lifshitz (1960), § 62). Namely we can write the integral for ω in the form

$$\bar{\omega} = \int_0^\infty \mathrm{Im}\, \frac{\epsilon_1(\omega) - 1}{\epsilon_1(\omega) + 1}\, d\omega = \int_0^\infty \frac{2\epsilon_1''(\omega) d\omega}{[\epsilon_1'(\omega) + 1]^2 + [\epsilon_1''(\omega)]^2} \quad . \quad (5.14)$$

where $\epsilon'(\omega)$ and $\epsilon''(\omega)$ are the real and imaginary parts of the dielectric permeabilities for real frequencies, i.e. experimentally directly measurable quantities.

Thus, for the actually observed thicknesses of helium film one would expect μ to vary as l^{-3}, and correspondingly the form of the film profile to be $l \sim z^{-1/3}$. Calculation of the coefficient in this dependence needs, however, a knowledge of the optical properties of the solid body (the wall) over a wide range of frequencies. We emphasize that the calculation of this coefficient on the basis of data on the interaction of individual helium atoms with a solid wall is in any case inadmissible.

We also give the expression for μ for 'large' film thicknesses ($l \gg \lambda_0$). The corresponding transform from (5.11) is made by introducing the variable $x = 2p\xi l/c$ instead of ξ and by replacing ϵ_1 by ϵ_{10}. The integrations over both dx and dp are carried out analytically, and as a result we have

$$\mu(l) = -\frac{3\hbar c(\epsilon_{30} - 1)}{32\pi^2 l^4} \frac{\epsilon_{10} - 1}{\epsilon_{10} + 1} \phi_{ad}(\epsilon_{10}) \quad . \quad . \quad . \quad . \quad (5.15)$$

with the function ϕ_{ad} taken from (4.38). For a metal $\epsilon_{10} \to \infty$ and $\psi_{ad} = 1$. For quartz, which is transparent from $\sim 0{\cdot}15\,\mu$ to several μ, it is also meaningful to consider the case when l lies within this region. The corresponding dependence $\mu(l)$ is found from the same formula (5.15), but ϵ_{10} does not refer to the electrostatic dielectric constant but to the optical value ϵ_1, i.e. the square of the refractive index in the transparent optical region (cf. note on page (195).

Equations (5.11), (5.12), (5.15) do not contain the temperature, i.e., strictly speaking they refer to absolute zero. However, the corrections for temperature must be relatively small, and there are no grounds for expecting any material change in the form of the profile of the film as the temperature is changed, either below or above the λ-point (outside its immediate neighbourhood).

The difficulties of experimental observation of the thickness and profile of a helium film under conditions which are sufficiently near to the ideal of thermal equilibrium are very great, and it is only very recently that they have been overcome to the extent that the data (in the helium II region) could be considered at all reliable (see the reviews by Jackson and Grimes (1958) and by Atkins (1957)).

In accordance with what was said in § 5.2, there is no physical basis for expecting a film profile of the form

$$\rho gz \sim al^{-3} + bl^{-2}.$$

Anderson *et al.* (1960) point out that their data on the thickness of a helium film on a steel surface (up to a height of 40 cm) are well described by a law of the form $\rho gz = al^{-3}$ (the absolute values of the thicknesses are not given).

The same law accounts for the results of the measurements by Ham and Jackson (1957) and by Grimes and Jackson (1959) (for heights of 0·4 to 7 cm) with a coefficient $a \approx 4{\cdot}5 \times 10^{-15}$ erg. A comparison of this value with the coefficient of l^{-3} in eqn. (5.12) (putting $\epsilon_{30} - 1 = 0{\cdot}057$) gives $\hbar\bar{\omega} \sim 7{\cdot}5$ eV. This is a reasonable value for a metal (steel).

The coefficients in (5.12) and (5.15) (for $\epsilon_{10} \to \infty$) are equal when $l = 3c/2\bar{\omega}$, i.e. in this case for $l \sim 5 \times 10^{-6}$ cm. This means that the experimentally observed film thicknesses (100–400 Å) are near the transition region between the l^{-3} and l^{-4} dependences.

Acknowledgment

In conclusion we would like to express our sincere gratitude to Professor L. D. Landau for numerous discussions of questions considered in this review.

References

Abrikosov, I. I., 1957, *J. exp. theor. Phys.*, **33**, 799 (*Soviet Physics, JETP*, **6**, 615, 1958).

Abrikosov, A. A., Gorkov, L. P., and Dzyaloshinskii, I. E., 1959, *J. exp. theor. Phys.*, **36**, 900 (*Soviet Physics, JETP*, **9**, 636, 1959).

Anderson, O. T., Liebenberg, D. H., and Dillinger, J. R., 1960, *Phys. Rev.*, **117**, 39.

Atkins, K. R., 1954, *Canad. J. Phys.*, **32**, 347; 1957, *Progress in Low Temperature Physics*, Vol. 2 (Amsterdam).

Casimir, H. B. C., 1948, *Proc. Acad. Sci. Amst.*, **60**, 793.

Casimir, H. B. C., and Polder, D., 1948, *Phys. Rev.*, **73**, 360.

Deryagin, B. V., and Abrikosova, I. I., 1956, *J. exp. theor. Phys.*, **30**, 993; **31**, 3 (*Soviet Physics, JETP*, **3**, 819, 1957; **4**, 2, 1957).

Deryagin, B. V., Abrikosova, I. I., and Lifshitz, E. M., 1956, *Quart. Rev. chem. Soc.*, **10**, 295; 1958, *Uspekhi Fizich. Nauk*, **64**, 493.

Dzyaloshinskii, I. E., 1956, *J. exp. theor. Phys.*, **30**, 1152 (*Soviet Physics, JETP*, **3**, 977, 1957).

Dzyaloshinskii, I. E., Lifshitz, E. M., and Pitaevskii, L. P., 1959, *J. exp. theor. Phys.*, **37**, 229 (*Soviet Physics, JETP*, **10**, 161, 1960).

Dzyaloshinskii, I. E., and Pitaevskii, L. P., 1959, *J. exp. theor. Phys.*, **36**, 1797 (*Soviet Physics, JETP*, **9**, 1282, 1959).

Fradkin, E. S., 1959, *J. exp. theor. Phys.*, **36**, 1286 (*Soviet Physics, JETP*, **9**, 912, 1959).

Franchetti, S., 1957, *Nuovo Cim.*, **5**, 183.

Ginsburg, V. L., and Pitaevskii, L. P., 1958, *J. exp. theor. Phys.*, **34**, 1240 (*Soviet Physics, JETP*, **7**, 858, 1958).

Grimes, L. G., and Jackson, L. C., 1959, *Phil. Mag.*, **4**, 1346.

Ham, A. C., and Jackson, L. C., 1957, *Proc. roy. Soc.* A, **240**, 243.

Hamaker, H., 1937, *Physica*, **4**, 1058.

Jackson, L. C., and Grimes, L. G., 1958, *Advanc. Phys.*, **7**, 435.

De Jongh, J. G. V., 1958, Dissertation, Urecht.

Kitchener, J. A., and Prosser, A. P., 1957, *Proc. roy. Soc.* A, **242**, 403.

Landau, L. D., 1958, *J. exp. theor. Phys.*, **34**, 262 (*Soviet Physics, JETP*, **7**, 182, 1958).

Landau, L. D., and Lifshitz, E. M., 1958, *Statistical Physics* (London: Pergamon Press) 1959, *Fluid Mechanics* (London: Pergamon Press); 1960, *The Electrodynamics of Continuous Media* (London: Pergamon Press).

Lifshitz, E. M., 1955, *J. exp. theor. Phys.*, **29**, 94 (*Soviet Physics, JETP*, **2**, 73, 1956).

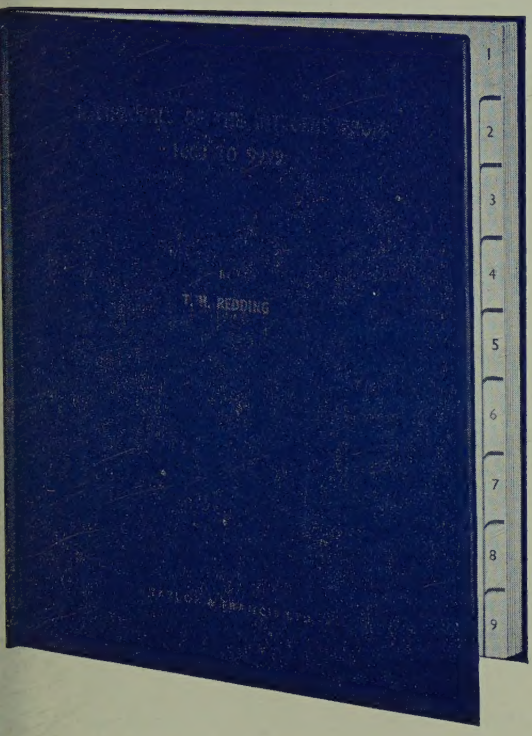
London, F., 1930, *Z. Phys.*, **60**, 491.

Matsubara, T., 1955, *Progr. theor. Phys.*, **14**, 351.

Mott, N. F., 1949, *Phil. Mag.*, **40**, 61.

Pitaevskii, L. P., 1959, *J. exp. theor. Phys.*, **37**, 577 (*Soviet Physics, JETP*, **10**, 408, 1959).

FOR USERS OF CALCULATORS



RECIPROCALS OF THE INTEGERS FROM 1000 TO 9999

By T. H. Redding
M.Sc.(Lond.), A.M.I.Mech.E., A.F.R.Ae.S.

Arranged for use with mechanical calculating machines to facilitate the evaluation of quotients and compound fractions

With an Appendix on mechanical barrel-setting calculators

These tables list the 10^6 multiples of the reciprocals of the integers 1000 to 9999 and are displayed as 1000 entries on each of nine double-page spreads which are thumb-indexed (1000 . . . , 2000 . . . , 3000 . . . , . . . etc) to facilitate rapid manipulation with one hand only. The entries are direct reading and are correct to the nearest integer in the third place of decimals.

Although suitable for general use, the tables have been prepared and arranged in a manner particularly suitable for use in conjunction with mechanical computing machines of the "barrel-setting" type in the evaluation of quotients and compound fractions. Equally the tables find application in effecting the evaluation of quotients on adding and listing machines when these incorporate a mechanism for automatically evaluating simple products.

In the first instance it was anticipated that an Appendix dealing with the advantages to be gained by the use of the reciprocal function in conjunction with such machines—and particularly with barrel-setting machines—would suffice to ensure the best use being made of the tables. The present 22-page *Appendix*, however, goes further than this, since the discussion of the tables is prefaced by a description of barrel-setting machines and their (simplest) method of operation in the evaluation of products, quotients and compound fractions. Thus, the description of each calculating procedure is followed by a numerical example with accompanying diagrams to illustrate the keyboard-displays at the beginning, the end, and at intermediate stages in the calculation. The machine-evaluation of series comprising the sum of a number of product or quotient-terms is also discussed. Generally it is hoped that the Appendix will serve as an introduction to mechanical methods of computation and that it will materially assist prospective purchasers in the choice of a machine.

A valuable guide to all users of Computers. Bound in blue publishers case, lettered on the front, with an 18 page table index. Printed on good quality paper for constant use.

Size 10 in. × 8 in.

Price 18s. 6d. *plus postage and packing* 1s 3d.

Printed and Published by

TAYLOR & FRANCIS LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

FILED IN STACKS

Announcing the new publication

INSTRUMENT CONSTRUCTION

Translated from the Russian

Приборостроение

Editor-in-chief: M. E. Rakovskii. Sub-editor: Yu. I. Shendler

The Russians describe their publication as a ' scientific, technical and production ' journal. It covers industrial instruments and instrumentation, automatic control, and production engineering for precision work. The articles which it presents not only introduce new instruments and techniques, they also afford a valuable insight into current Russian practice.

Subscription £6 yearly post free ($17.10 U.S.A. and Canada). A special rate of £3 yearly post free ($8.55 U.S.A. and Canada) is available to University and Technical College libraries.

Single copies 15*s*. each ($2.15 U.S.A. and Canada) plus postage

Orders should be sent to the Subscription Department, Taylor & Francis Ltd.

Produced for the Department of Scientific and Industrial Research by the British Scientific Instrument Research Association, ' Sira ', South Hill, Chislehurst, Kent

Printed and Published for B.S.I.R.A. by

TAYLOR & FRANCIS LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4